Preparing for First Beam at the LHC: An ALICE Perspective (The Tragedy of the Anti-Commons)

L. Pinsky

University of Houston

On behalf of the ALICE Computing Project

Gelato ICE

April 24-25, 2006

San Jose, California

ALICE-USA Collaboration

AliEn 2 @GRID

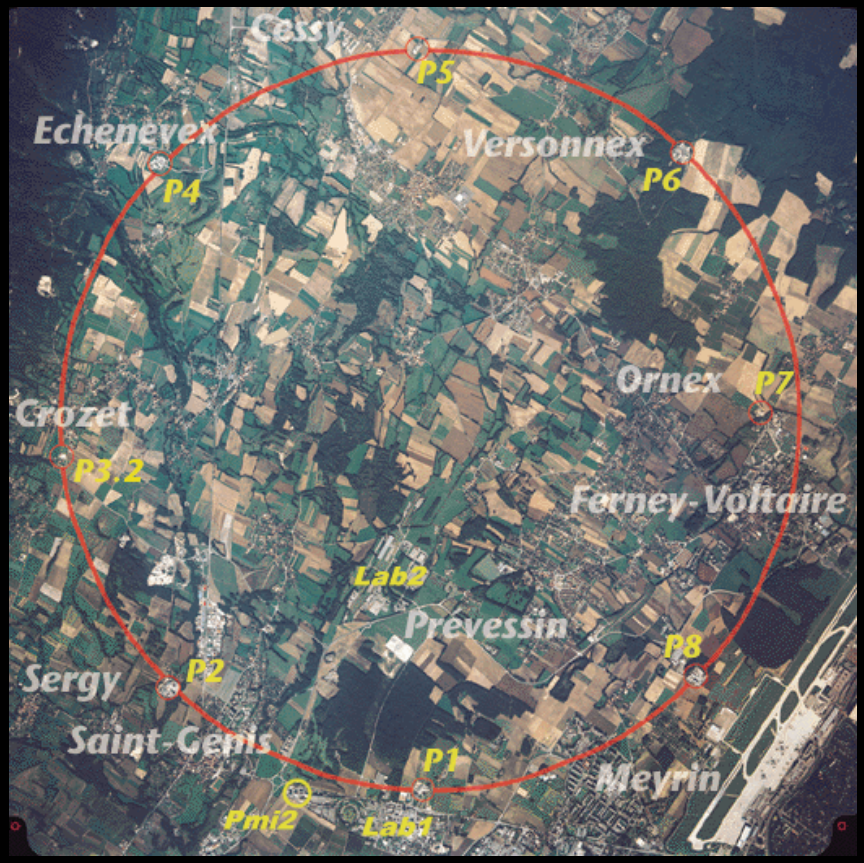
Roadmap…

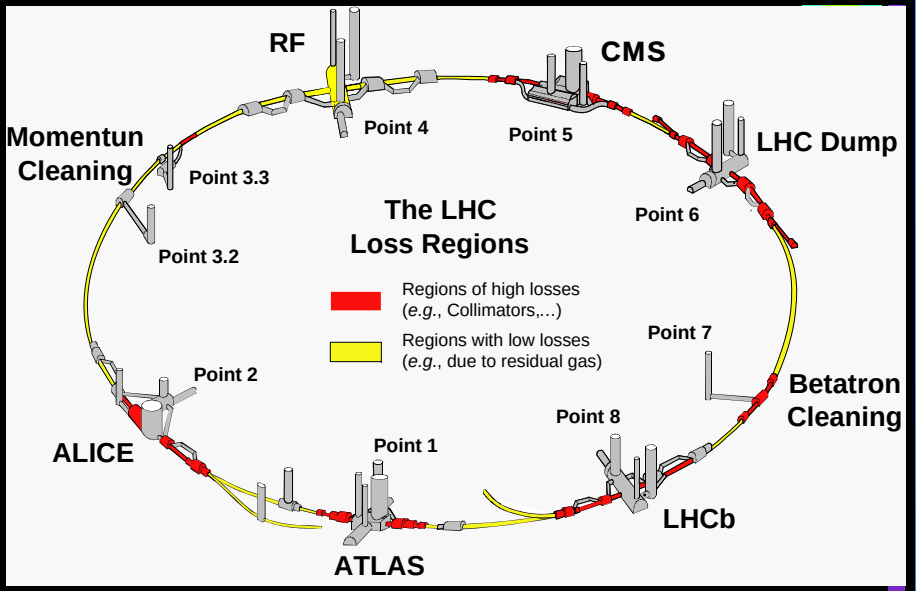
- Acknowledgments and Disclaimers…
- A little bit about LHC & ALICE
- ALICE Hardware requirements
- ALICE's Software infrastructure
- Testing the system (Data Challenges)
- Some results…
- Itanium relative performance
  - (warning… mostly not good news…)
- …but, there is at least some good news

Acknowledgments and Disclaimers

- I am an experimental particle physicist, AND I am an Intellectual Property Attorney...
  - (You are warned...)
- Much of the content of this talk comes from my colleagues at ALICE:
  - Federico Carminati & Latchezar Betev
- The Itanium angle is somewhat of a tragedy in the classical sense, sort of like my new MacBook Pro (Intel-based Apple Laptop)
  - The current situation is dismal, but for reasons not related to the native capability of the hardware
  - ...But in both cases, the future is not yet determined...

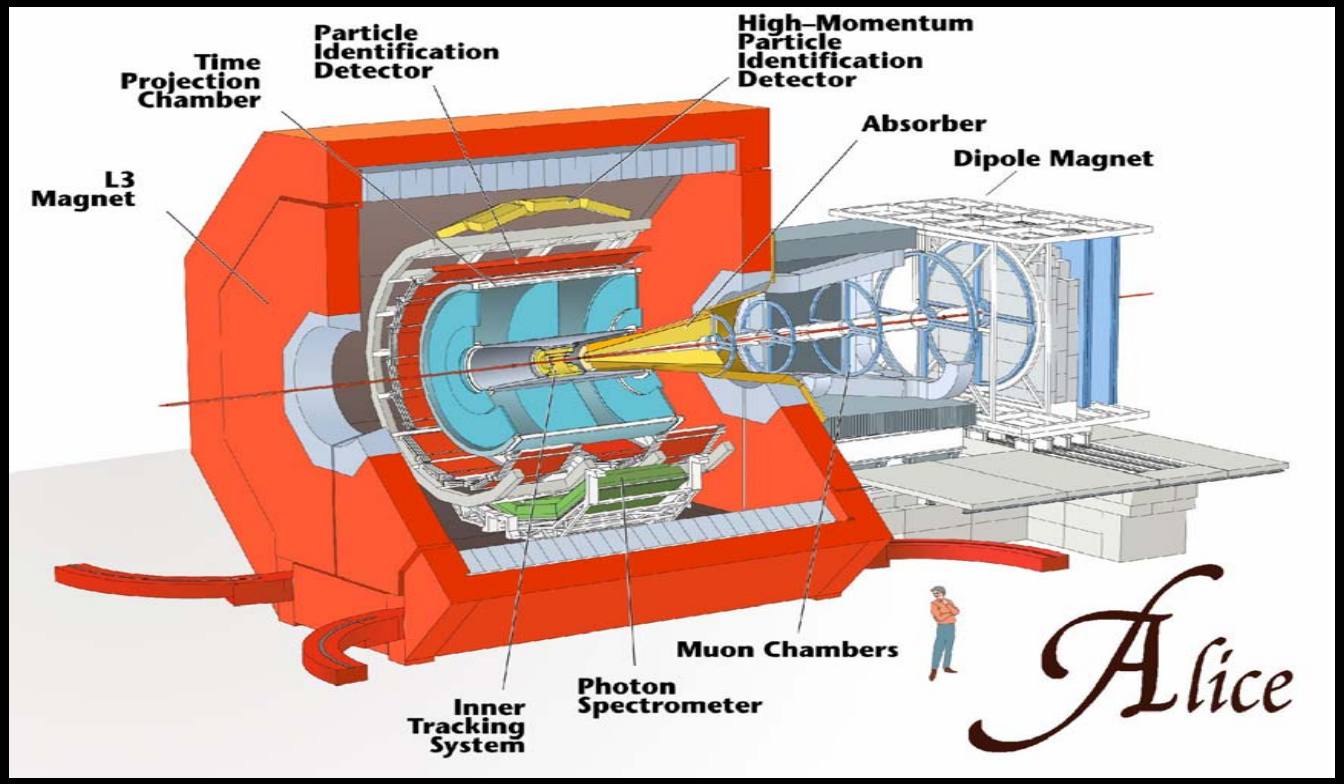
What is the LHC?
(Large Hadron Collider)





- CERN is in Geneva, Switzerland
- The LHC is 27km around!
- It will colide protons @ 14TeV
- ...& Pb at 5.5 TeV/nucleon
- Startup is set for next year...
- 4 Expts: ATLAS, CMS, LHCb & ***ALICE***...

What is ALICE (A Large Ion Collider Experiment)



Total weight	**10,000t**
Overall diameter	**16.00m**
Overall length	**25m**
Magnetic Field	**0.4Tesla**

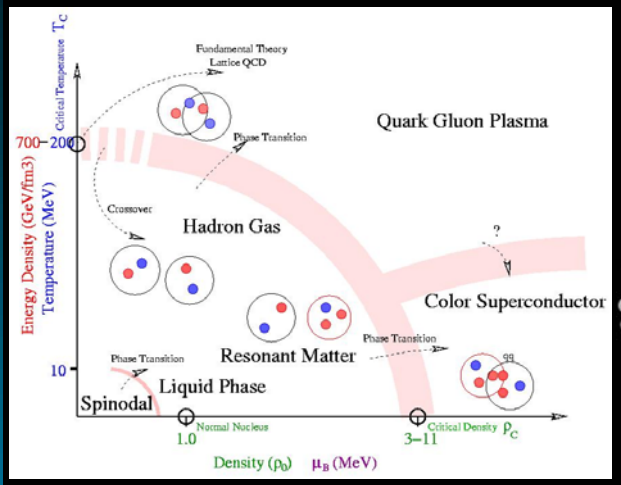
ALICE Collaboration

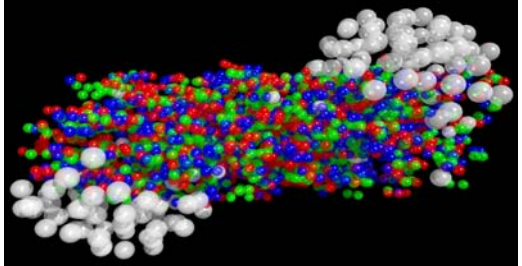
- ~ 1/2 ATLAS, CMS, ~ 2x LHCb
- ~1000 people, 30 countries, ~ 80 Institutes

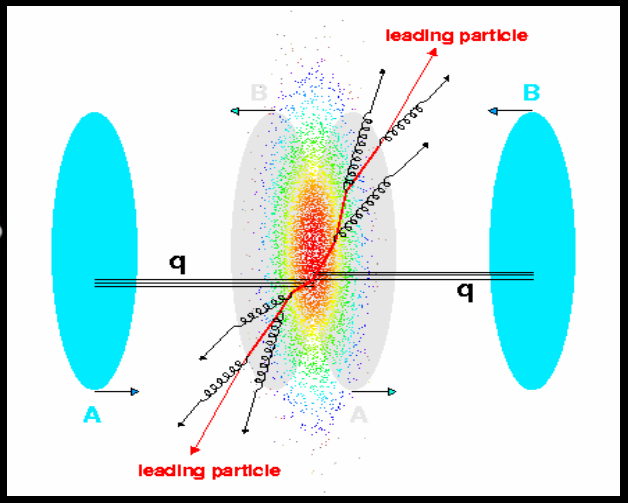
The First Element of the Tragedy

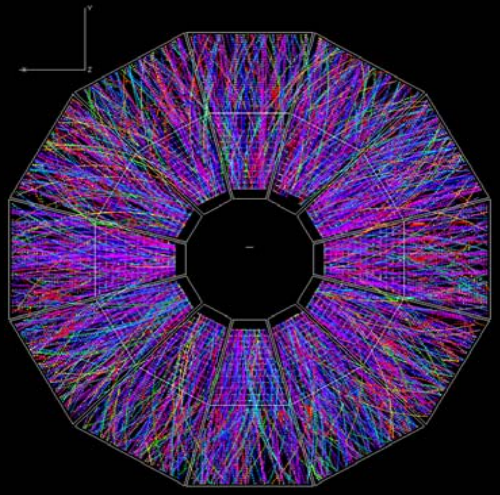
- CMS and ATLAS, the two experiments with the biggest clout, each have twice the number of participants as ALICE...
- However, the computing needs of ALICE are at least as great (if not greater)
- This means that we need at least double the computing resources per participant with respect to the highest visibility LHC Experiments
- Funding agencies have not responded kindly to that situation (so far)...
- So, ALICE has been forced to embrace a "Genuine" Grid approach (Why this has lead to a tragedy will become clear later)...

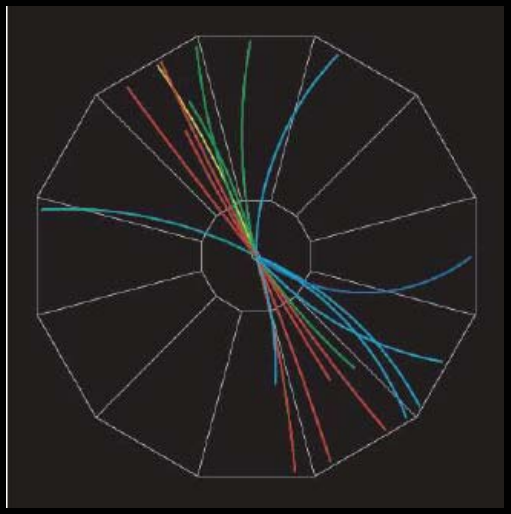
One Physics Slide (Sorry, but I'm a Physicist...)

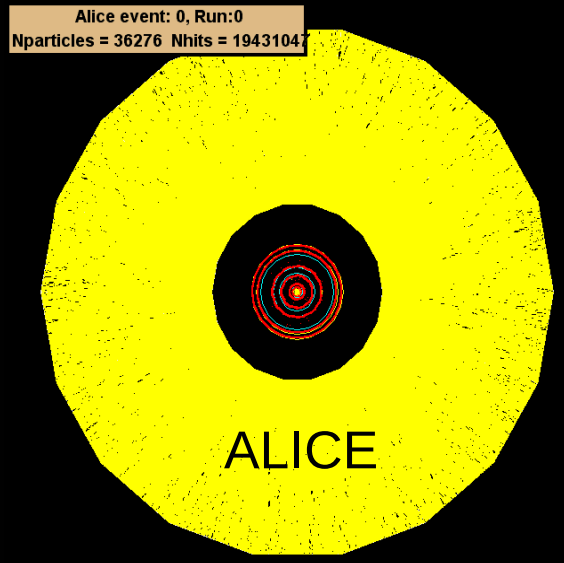










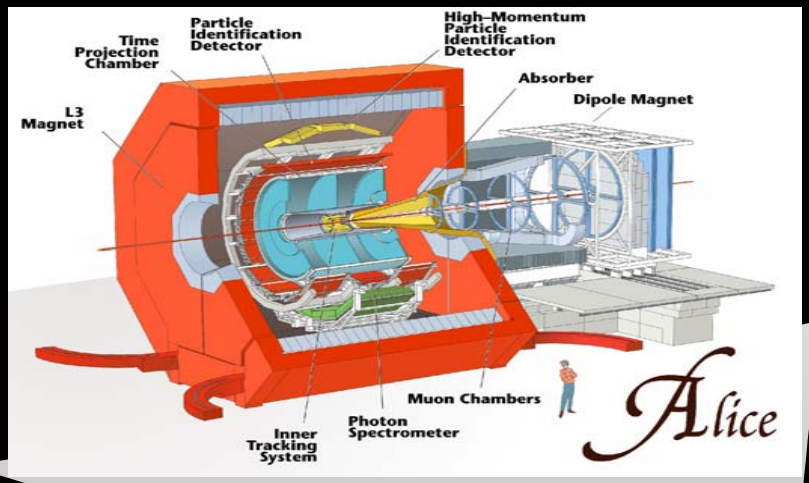


STAR @ RHIC

ALICE “Vital Statistics”

- One month of Pb-Pb (10^6 s) per year
- Raw rate ~4-8000 Hz
- Trigger rates (HZ) L0~400, L1~100, L2~10
- Pb-Pb Event size ~ 12-20 MB
- Archiving bandwidth ~1.25 GB/s ~10 Hz
- ...and also, 6-8 months/year of p-p data.
- The data taken each year must be fully analyzed before the next year’s data taking begins...

Data Flow



8 kHz (160 GB/sec)

level 0 - special hardware

400 Hz (4 GB/sec)

level 1 - embedded processors

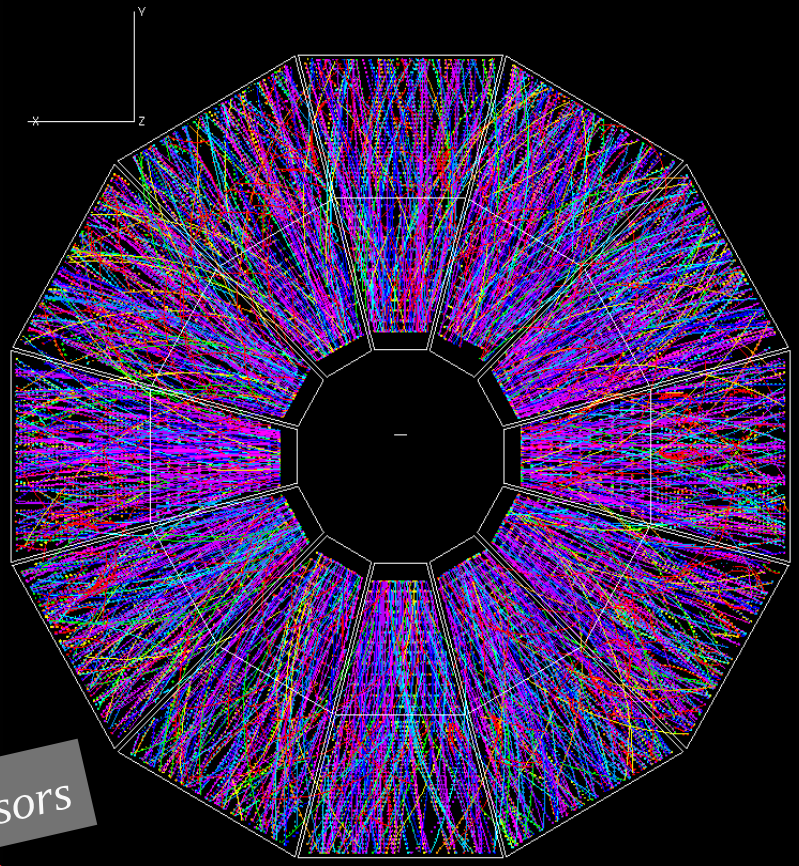
100 Hz (2.5 GB/sec)

level 2 - PCs

10 Hz

(1.25 GB/sec)

data recording & offline analysis



Actually from STAR @ RHIC
ALICE will be much worse
As already noted…

=> Distributed OffLine Grid Analysis

OffLine Analysis Infrastructure

- MONARC Model:
  - Tier 0 (CERN) -> 7 Tier 1's (National) ->
  - Multiple Tier 2's in support of each Tier 1.
  - Tier 1's have data archival responsiblilty
  - Data set is divided among the Tier 1's
  - Tier 2's provide CPU plus dynamic storage
- CLOUD Model:
  - ...Blurs the Tier 1-2 structure...
  - ALICE-USA is proposing a CLOUD structure

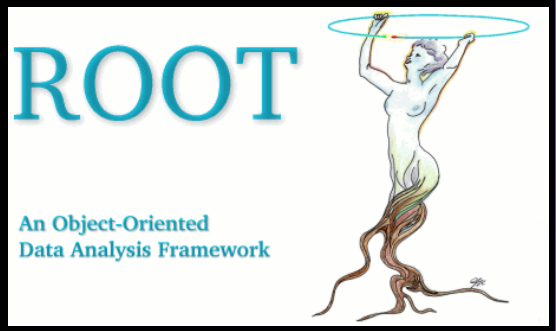
Summary of Global ALICE Computing Needs

Summary of Computing Capacities required by ALICE										
		Tier0	Tier1	Tier1ex	Tier2	Tier2ex	Total	at CERN	HR	LHCC review
CPU (MSI2K)	Maximum	7.5	13.8	13.8	13.7	13.7	35.0	7.5		34.0
		22%	39%	39%	39%	39%	100%	22%		
	Average	3.0	13.8	10.1	13.7	12.7	30.4	7.5	17.6	26.0
		10%	45%	33%	45%	42%	100%	25%		
DisK (Pbytes)		0.1	7.7	6.6	2.4	2.3	10.2	1.3	1.6	10.0
		1%	76%	65%	23%	22%	100%	13%		
MS (Pbytes/year)		2.3	7.5	6.4	-	-	9.8	3.3	4.7	10.7
		23%	77%	66%	-	-	100%	34%		
Network in (Gb/s)		8.00	2.00	-	0.01	-	-	-		
Network out (Gb/s)		6.00	1.50	-	0.27	-	-	-		

Summary of Computing Capacities pledged for ALICE							
	2005	2006	2007	2008	2009	2010	Comments
				Tier1			
CPU (MSI2K)	0.58	1.17	3.03	9.59	16.65	16.37	
DisK (Pbytes)	0.09	0.56	1.52	3.34	7.52	7.67	
MS (Pbytes)	0.17	0.97	2.63	5.92	11.60	9.98	
				Tier2			
CPU (MSI2K)	1.37	2.51	4.37	5.60	7.40	7.74	
DisK (Pbytes)	0.18	0.46	0.90	1.30	1.76	2.28	

ALICE Software

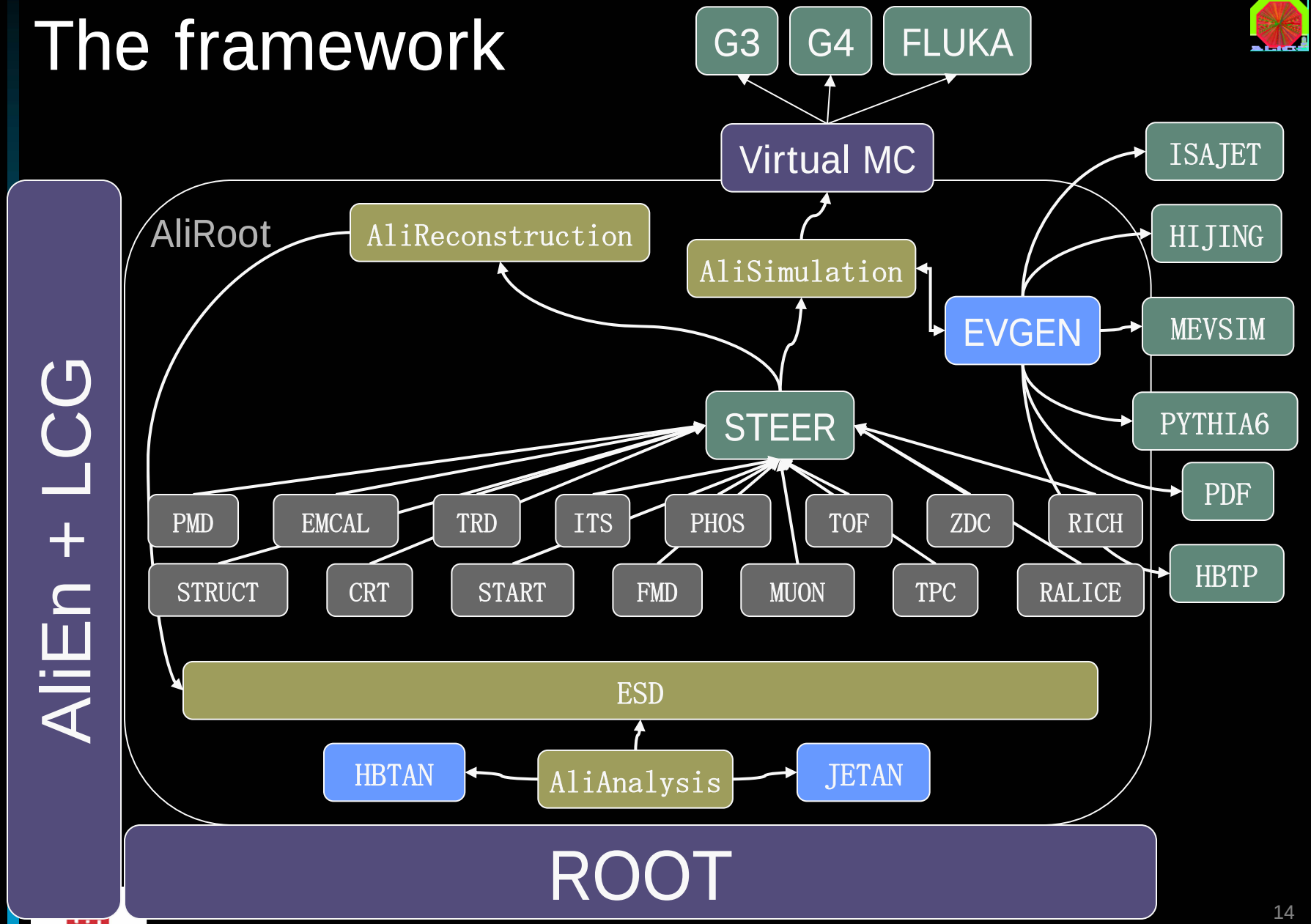
- ROOT - Object-Oriented Data Analysis Infrastructure (http://root.cern.ch/)
  - PROOF - Parallel ROOT
- AliEn - ALICE Environment Grid Middleware
- AliRoot - ROOT-based Grid analysis tool for ALICE based on AliEn...



The history

- Developed since 1998 along a coherent line
- Developed in close collaboration with the ROOT team
- No separate physics and computing team
  - Minimise communication problems
  - May lead to “double counting” of people
- Used for the TDR’s (Technical Design Reports) of all detectors and Computing TDR simulations and reconstructions

The framework



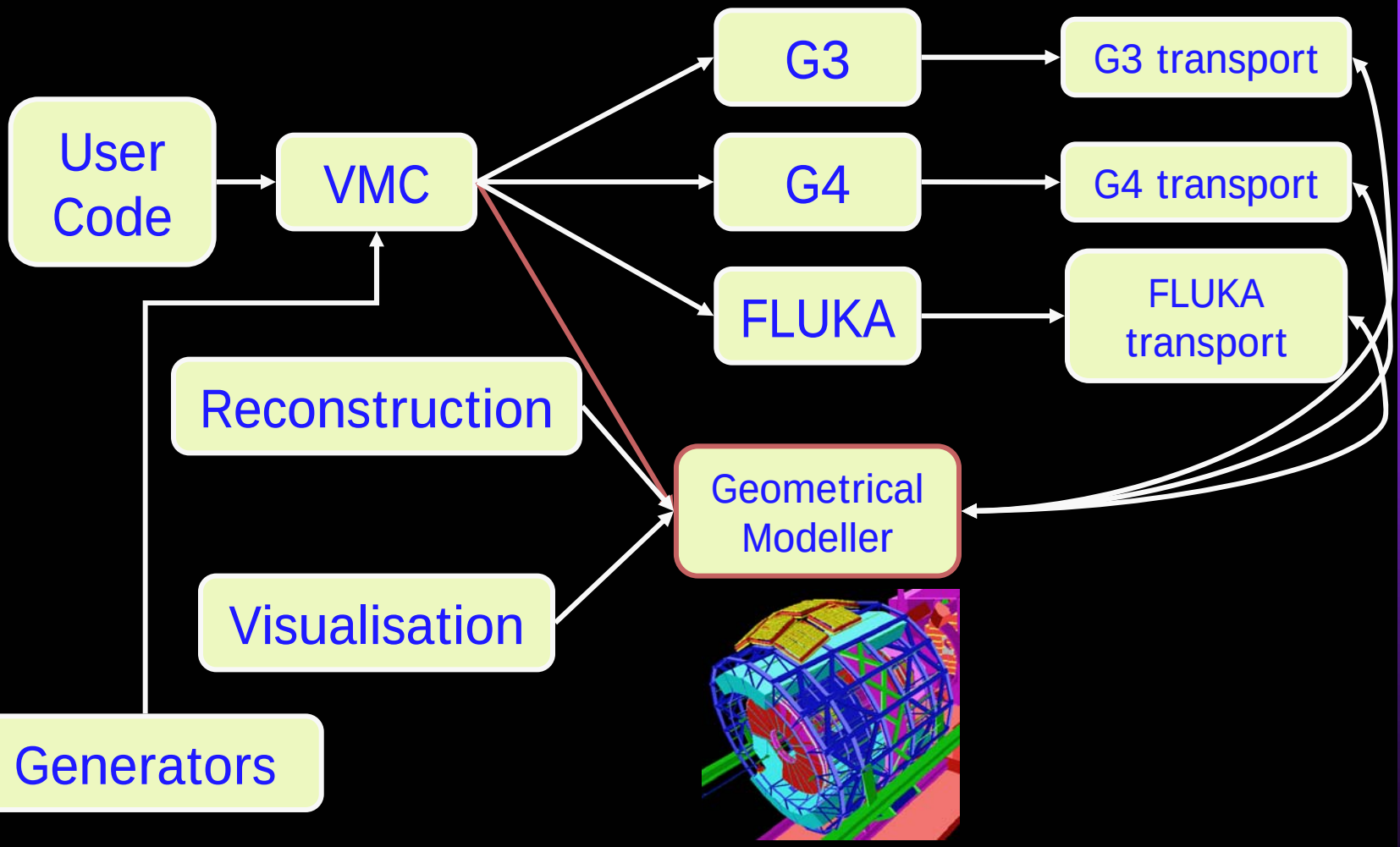
The code

- 0.5MLOC C++
- 0.5MLOC “vintage” FORTRAN code
- Nightly builds
- Strict coding conventions
- Subset of C++ (no templates, STL or exceptions!)
  - “Simple” C++, fast compilation and link (see R.Brun’s talk)
  - No configuration management tools (only cvs)
  - aliroot is a single package to install
- Maintained on several systems
  - DEC-Tru64, Mac OSX, Linux RH/SLC/Fedora (i32:i64:AMD), Sun Solaris
- 30% developed at CERN and 70% outside

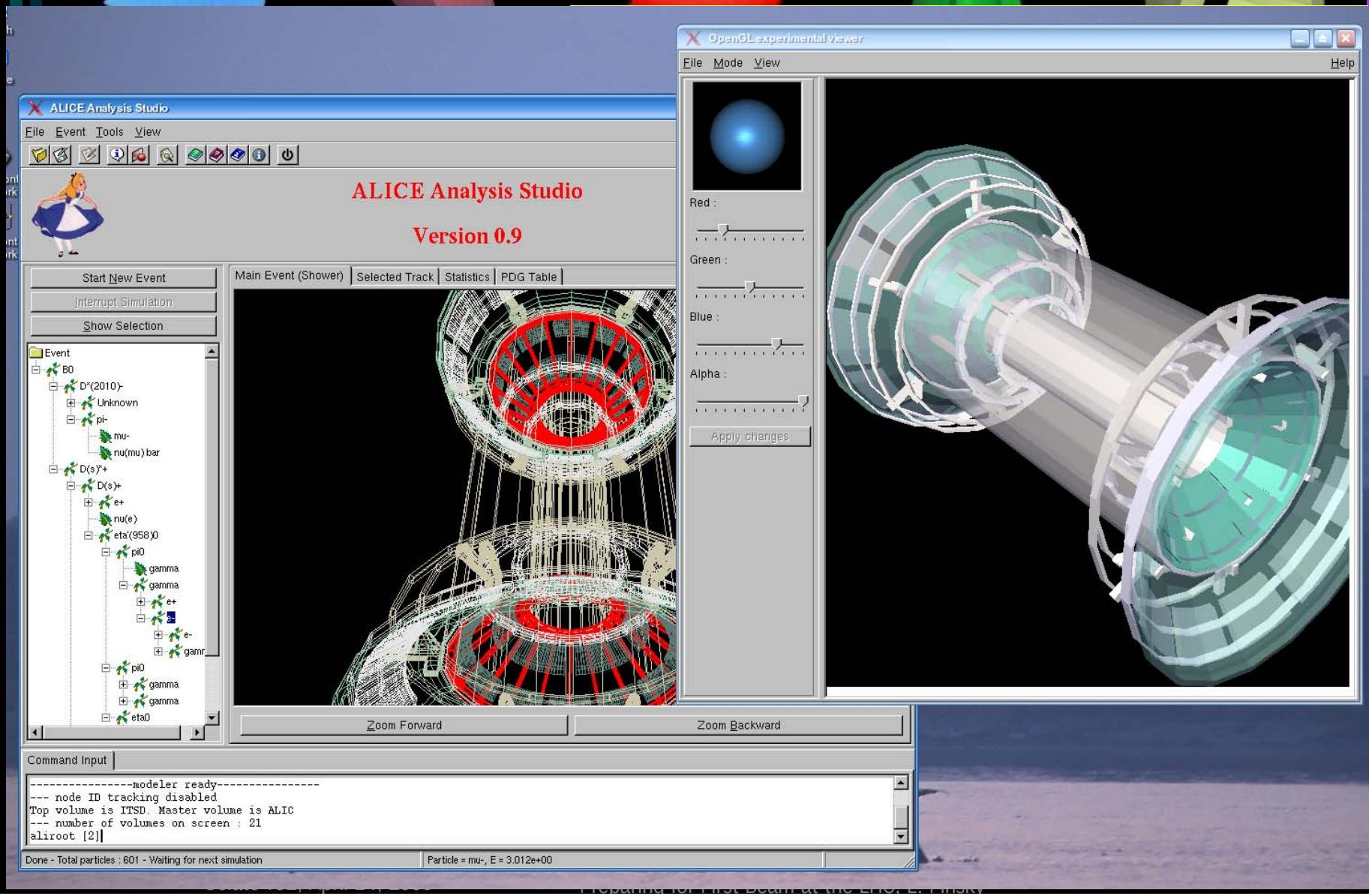
The tools

- Coding convention checker
- Reverse engineering
- Smell detection
- Branch instrumentation
- Genetic testing (in preparation)
- Aspect Oriented Programming (in preparation)

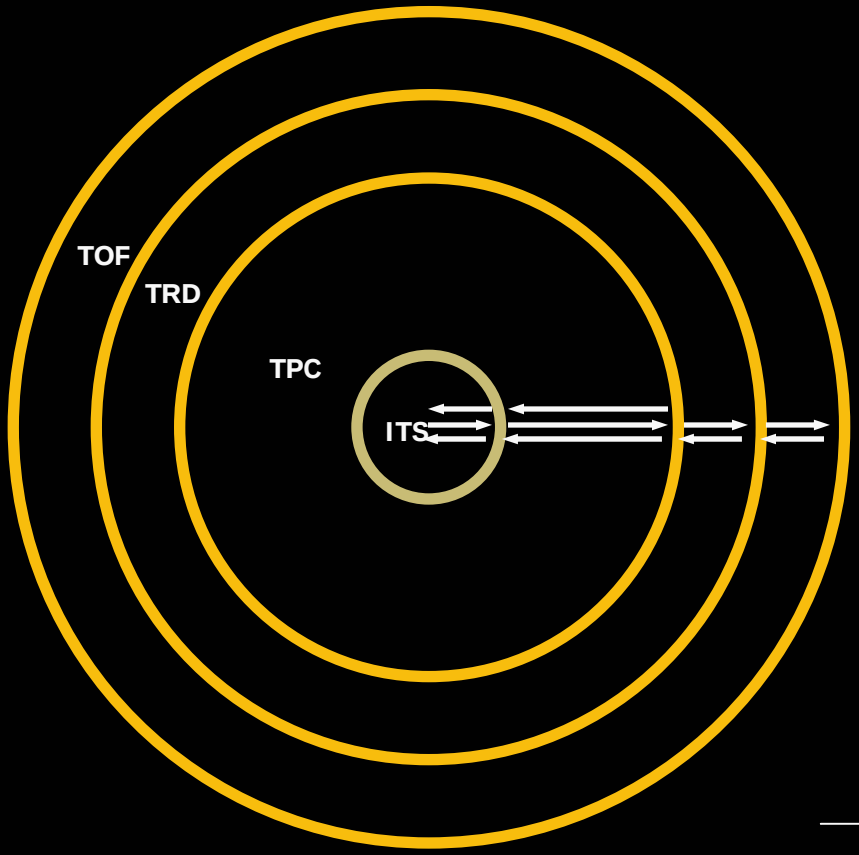
The Simulation - Virtual Monte Carlo



TGeo modeller

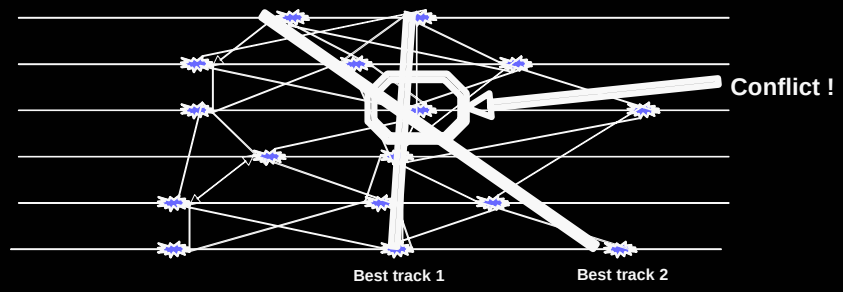


The reconstruction

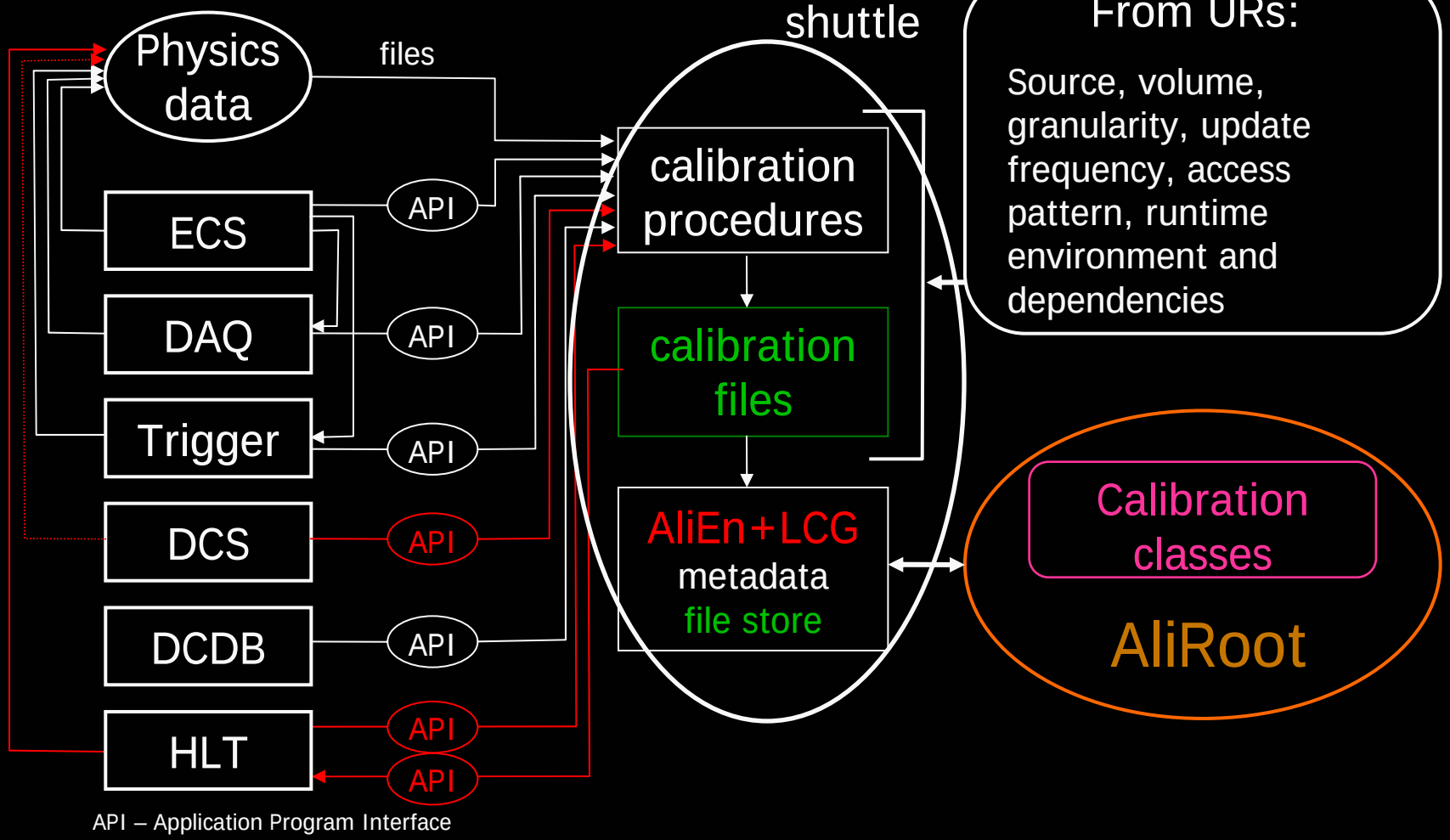


- Incremental process
  - Forward propagation towards to the vertex TPC⇒ITS
  - Back propagation ITS⇒TPC⇒TRD⇒TOF
  - Refit inward TOF⇒TRD⇒TPC⇒ITS
- Continuous seeding
  - Track segment finding in all detectors

- Combinatorial tracking in ITS
  - Weighted two-tracks χ^2 calculated
  - Effective probability of cluster sharing
  - Probability not to cross given layer for secondary particles

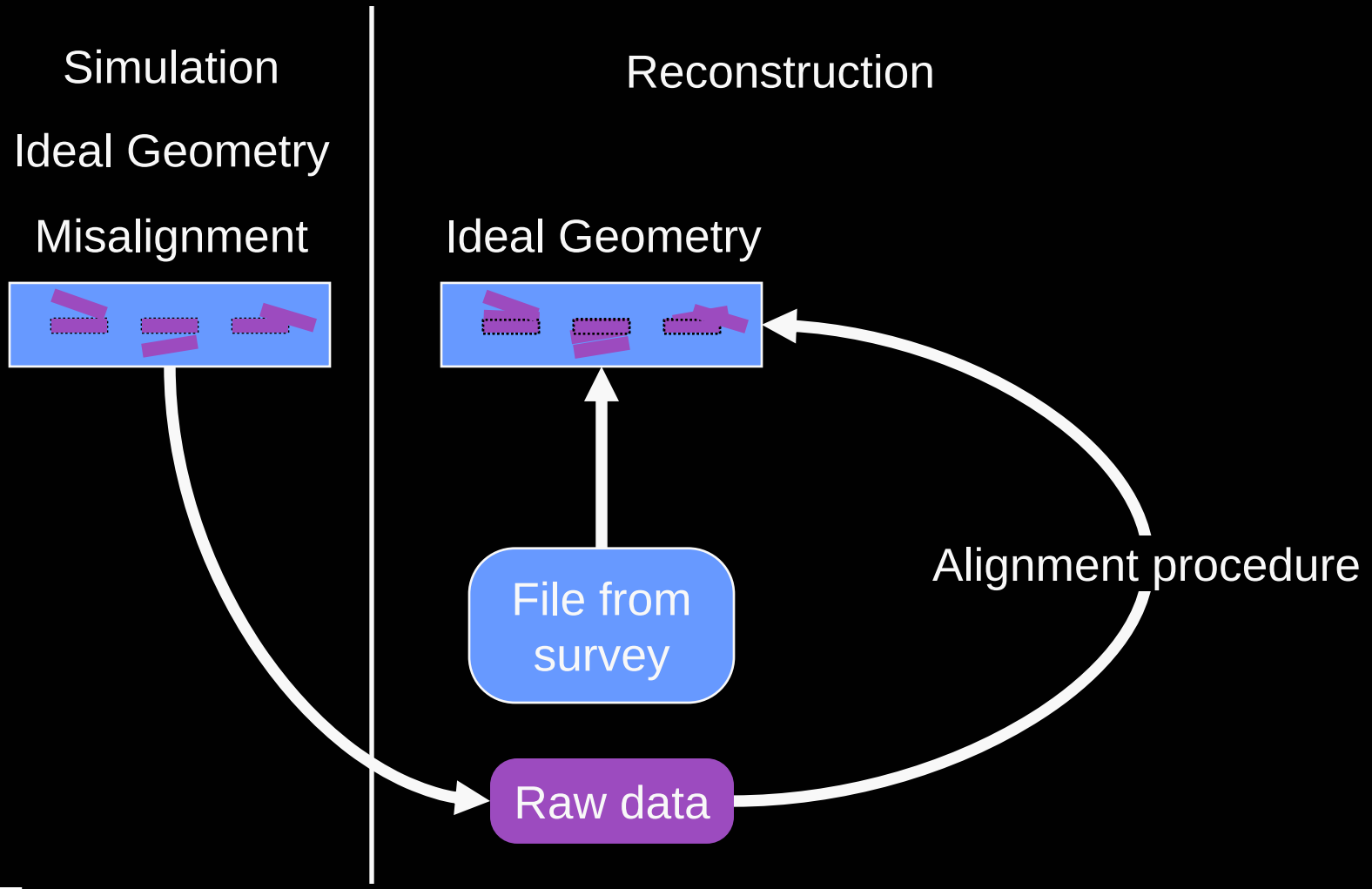


Calibration

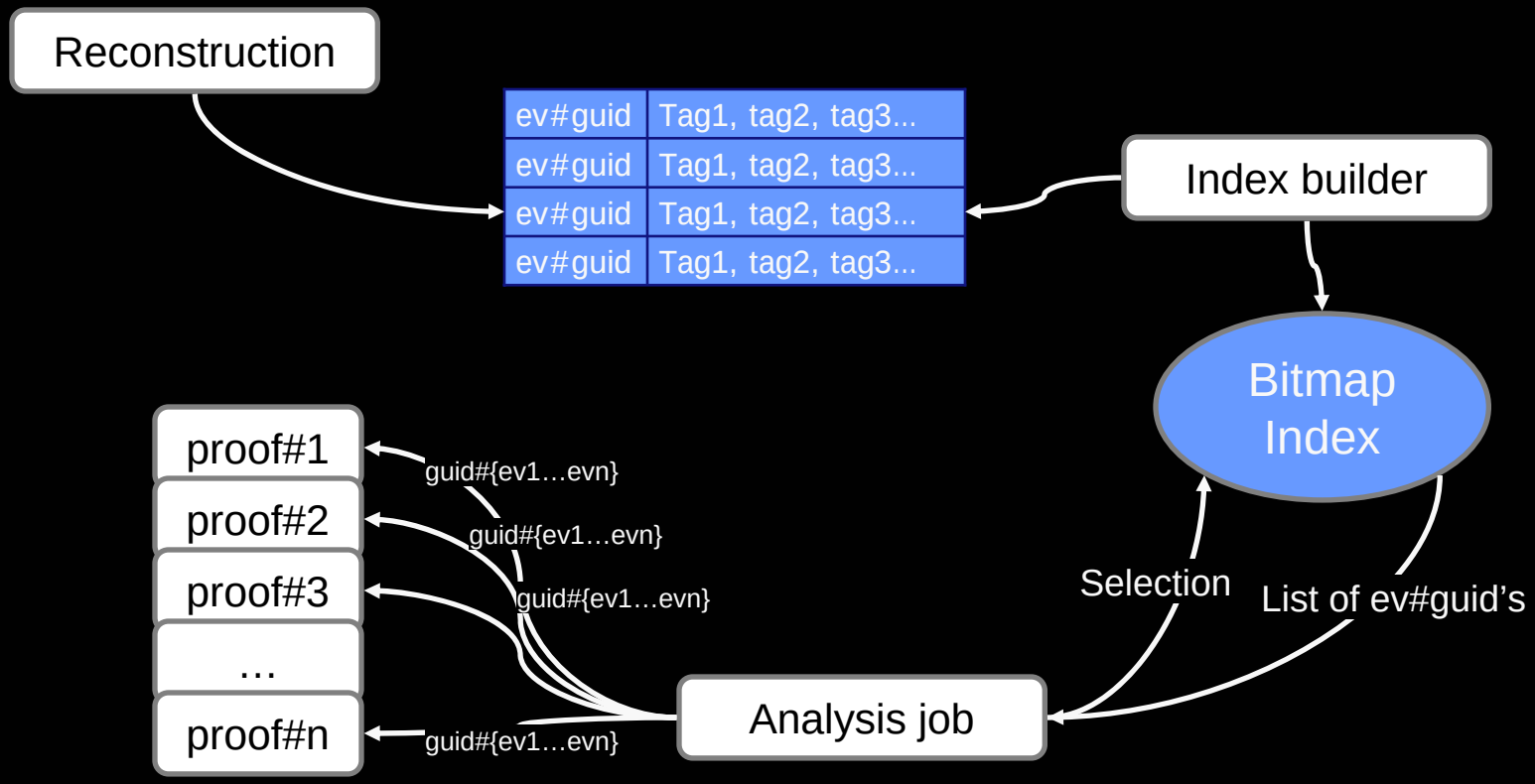


API – Application Program Interface

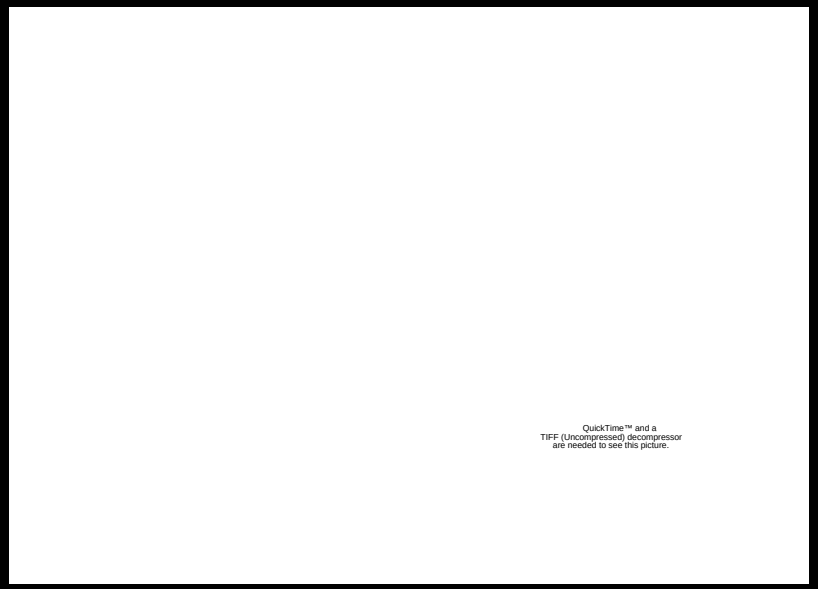
Alignment


Simulation
Ideal Geometry
Misalignment
Reconstruction
Ideal Geometry
File from survey
Raw data
Alignment procedure

Tag architecture

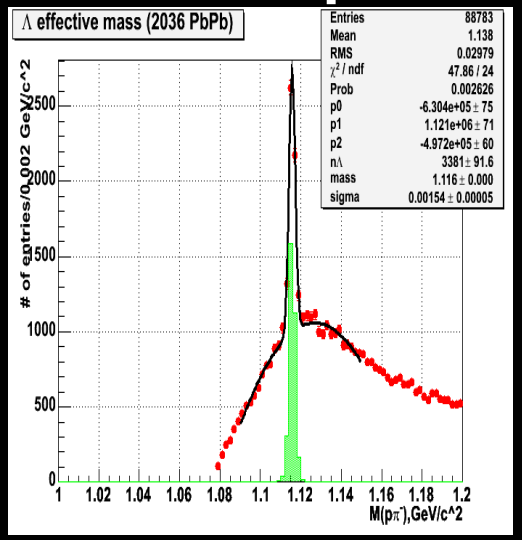


Visualisation

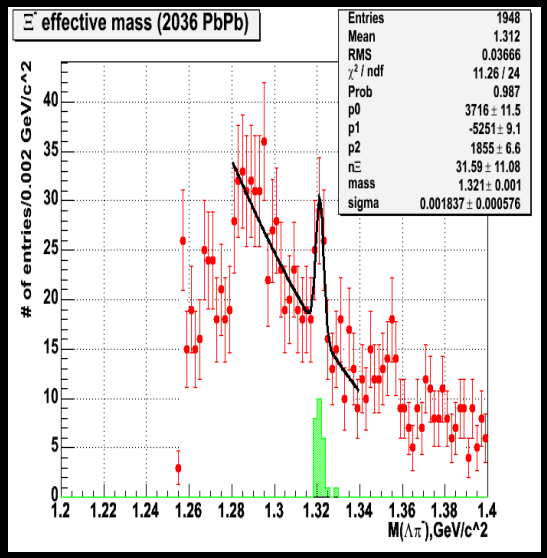


ALICE Analysis Basic Concepts

- Analysis Models
  - Prompt reco/analysis at T0 using PROOF infrastructure
  - Batch Analysis using GRID infrastructure
  - Interactive Analysis using PROOF(+GRID) infrastructure
- User Interface
  - ALICE User access any GRID Infrastructure via AliEn or ROOT/PROOF UIs
- AliEn
  - Native and "GRID on a GRID" (LCG/EGEE, ARC, OSG)
  - integrate as much as possible common components
    - LFC, FTS, WMS, MonALISA ...
- PROOF/ROOT
  - single + multitier static and dynamic PROOF cluster
  - GRID API class TGrid(virtual)⇒TAliEn(real)



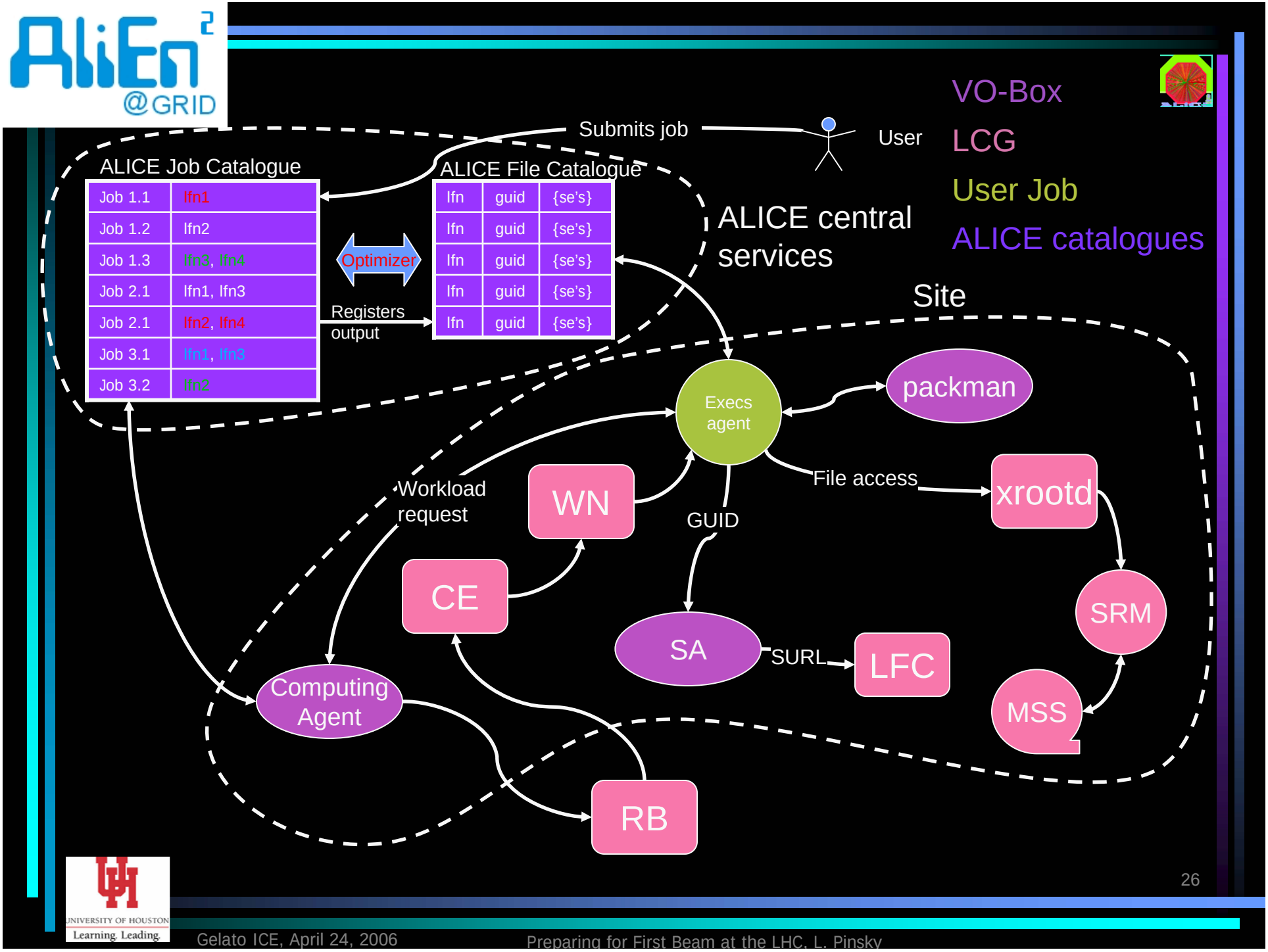
$\Xi \rightarrow \pi\Lambda \rightarrow p\pi$



ALICE view on the current situation

Exp specific services (AliEn' for ALICE)

EGEE, ARC, OSG…

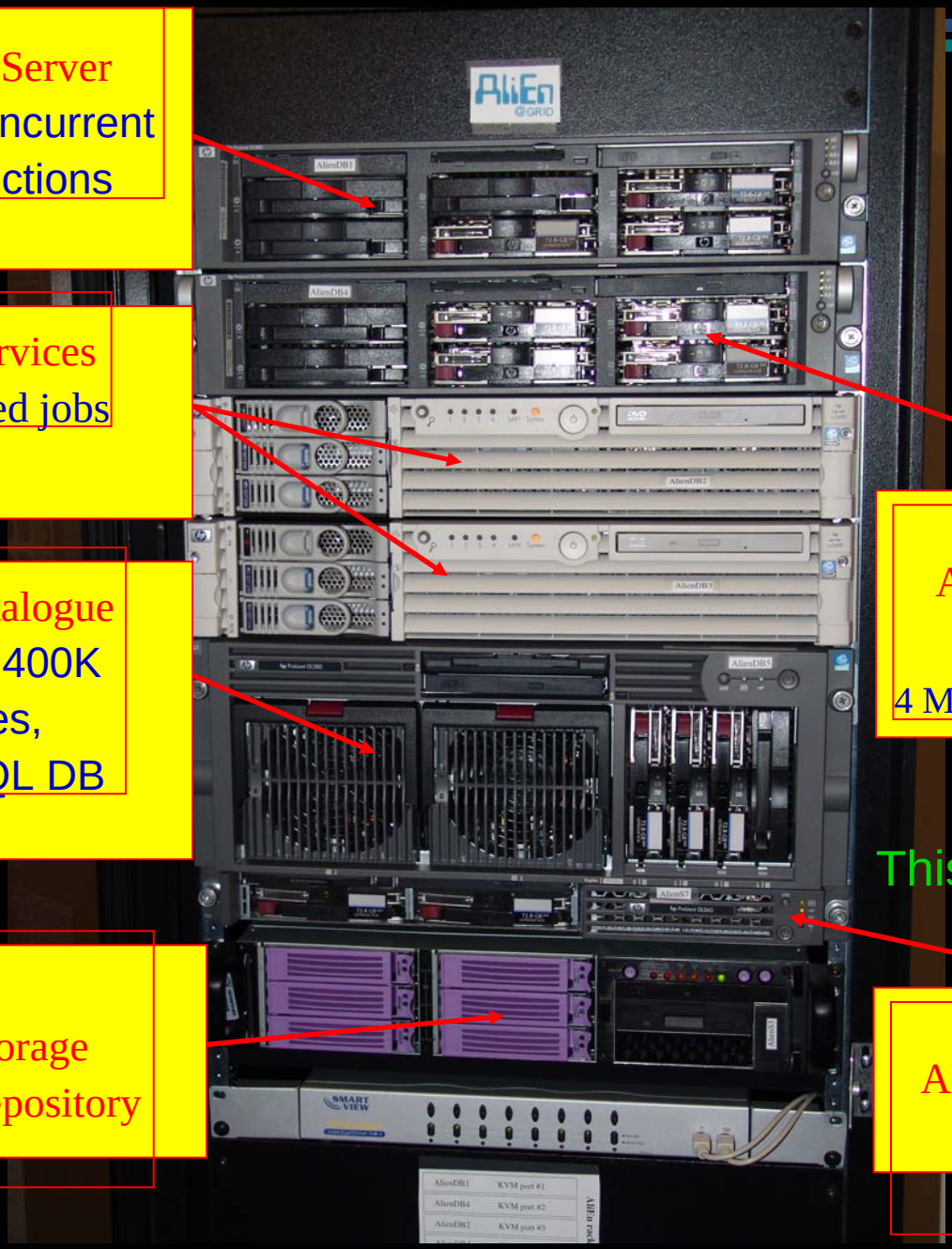

VO-Box
LCG
User Job
ALICE catalogues
Submits job
User
ALICE Job Catalogue
ALICE File Catalogue
Job 1.1 lfn1
Job 1.2 lfn2
Job 1.3 lfn3, lfn4
Job 2.1 lfn1, lfn3
Job 2.1 lfn2, lfn4
Job 3.1 lfn1, lfn3
Job 3.2 lfn2
lfn guid {se's}
Optimizer
Registers output
ALICE central services
Site
Execs agent
packman
Workload request
WN
File access
xrootd
GUID
CE
SRM
SA
SURL
LFC
Computing Agent
MSS
RB

AliEn Proxy Server
Up to 2000 concurrent client connections

AliEn Job Services
300 K archived jobs

AliEn File Catalogue
8Mio entries, 400K directories,
10GB MySQL DB

log files storage
MonALISA repository



AliEn Core services

AliEn Storage Elements
Volume Manager
4 Mio entries, 2GB MySQL DB

This is the Good News

AliEn to CASTOR (MSS) interface

Distributed analysis

File Catalogue query

User job (many events)

Data set (ESD's, AOD's)

Job output

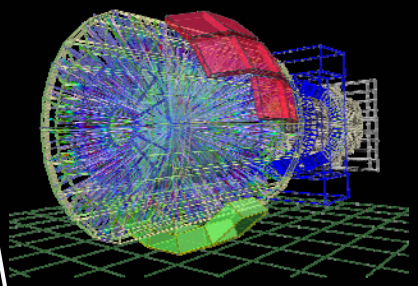
Job Optimizer

Sub-job 1

Sub-job 2

Sub-job n

Grouped by SE files location



Job Broker

Submit to CE with closest SE

CE and SE processing Output file 1

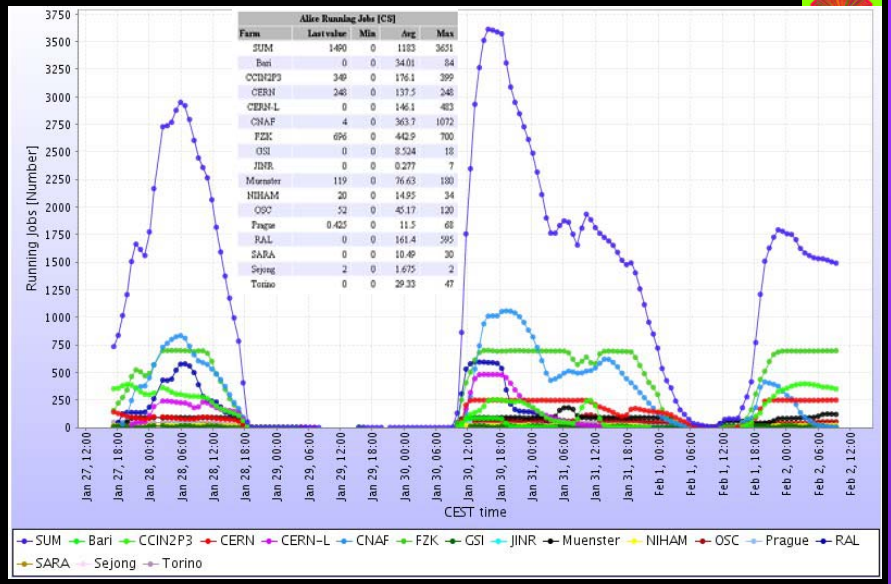
CE and SE processing Output file 2

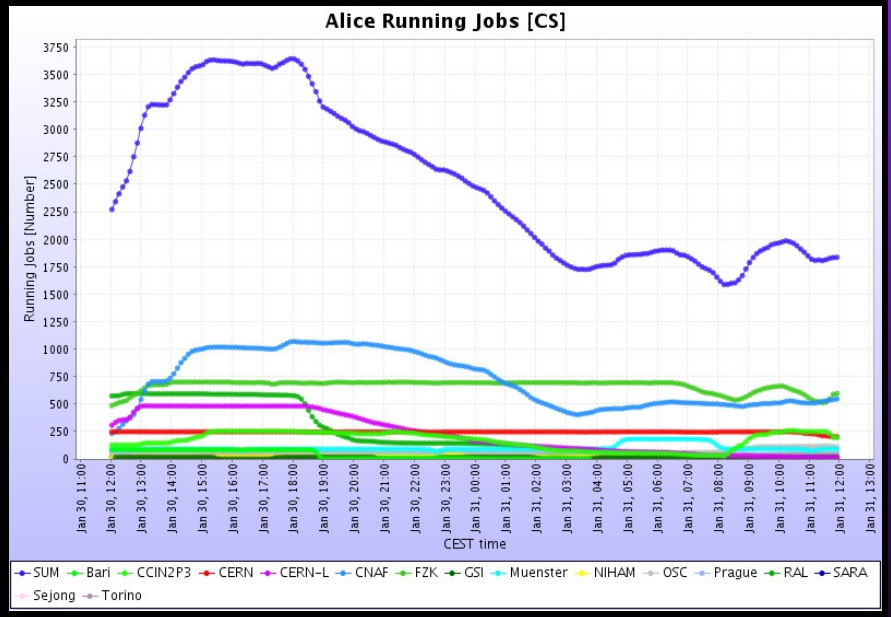
CE and SE processing Output file n

File merging job

Data Challenge

- Last (!) exercise before data taking
- Test of the system started with simulation
- Up to 3600 jobs running in parallel
- Next will be reconstruction and analysis





2004 DC - Principles and platforms

- True GRID data production and analysis: all jobs are run on the GRID, using only ***AliEn*** for access and control of native computing resources
- LCG GRID: access through AliEn-LCG inteface

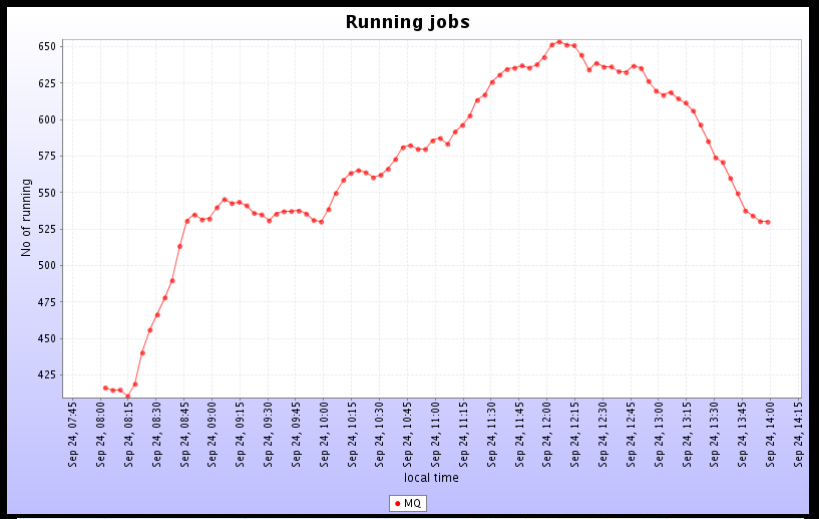
- In phase 3: gLite ***+PROOF (ARDA E2E Prototype for ALICE)***

- Software: AliRoot/GEANT3/ROOT/gcc3.2 libraries - distributed by AliEn
- Used platforms:
  - GCC 3.2 + i686 32-bit Cluster
  - GCC 3.2 + ia64 Itanium Cluster

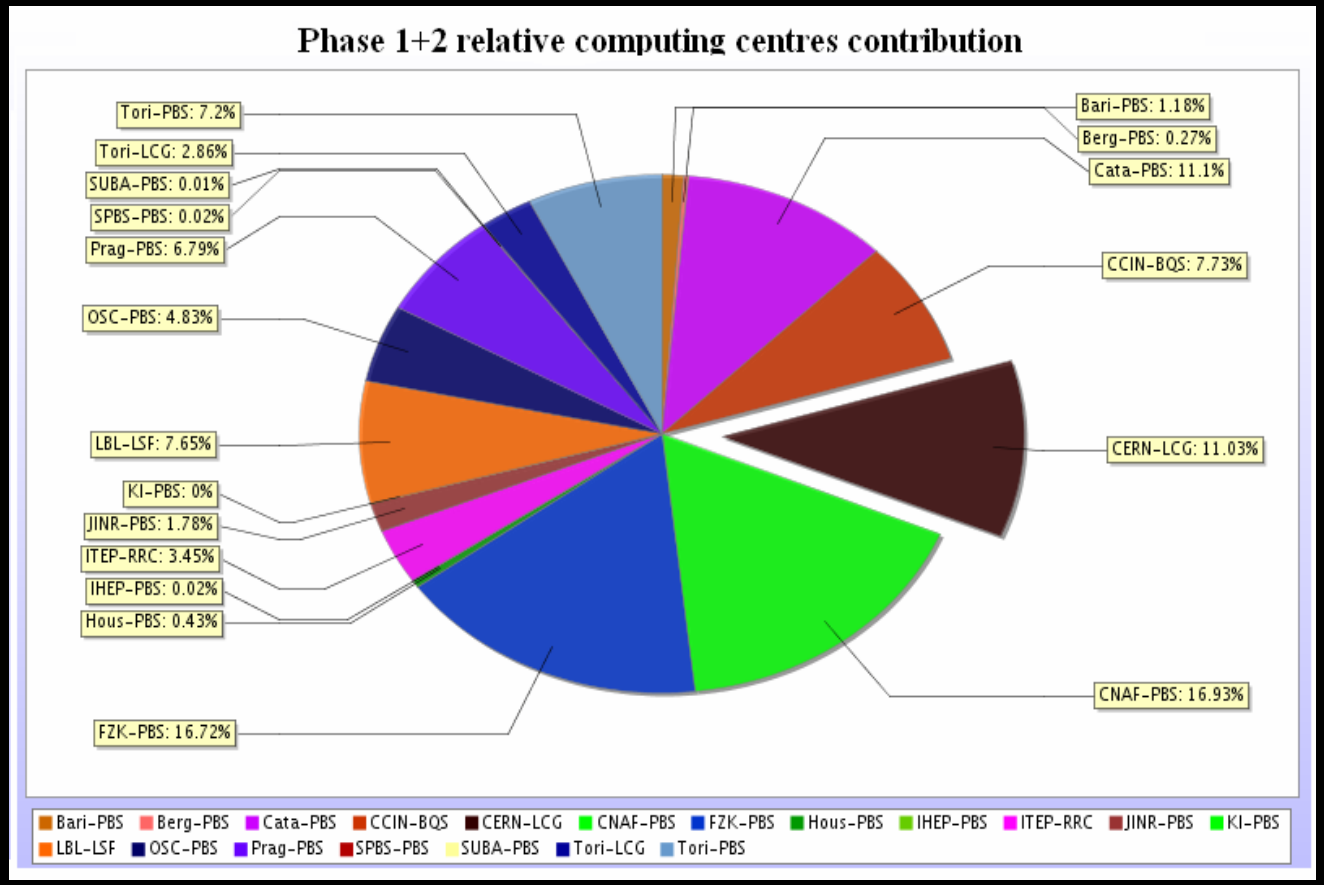
Monitoring

- ALICE repository – history of the entire DC
- ~ 1 000 monitored parameters:
  - Running, completed processes
  - Job status and error conditions
  - Network traffic
  - Site status, central services monitoring
  - ....
- 7 GB data
- 24 million records with 1 minute granularity

MonALISA
MONitoring Agents using a Large Integrated Services Architecture



2004 DC Site participation



- 17 sites under AliEn direct control and additional resources through GRID federation:
  - LCG sites accessed through an interface

Challenges'06

- Last Data Challenge before data taking
- Computing Data Challenge
  - Final version of rootifier / recorder
  - Online data monitoring
- Physics data challenge
  - Final version of simulation / reconstruction
  - Data analysis
- Proof data challenge
  - Preparation of the fast reconstruction / analysis framework under discussion
- Will simulate the realistic data flow with raw data
- Physics Working Groups are providing the requirements for physics
- Calibration / Alignment / Trigger
  - We are now able to mis-align/align and de-calibrate/calibrate for some detectors
  - We should have trigger masks for the relevant detectors

Services for SC4 (Current 2006 ongoing Data Challenge)

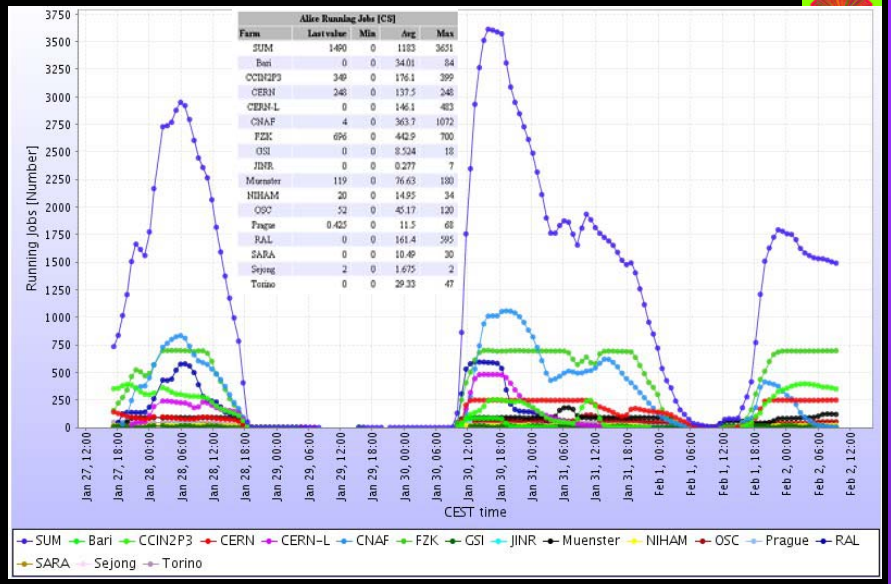
VO-box deployment on SC3 sites	ALICE
ROOT/AliRoot deployment on SC3 sites	ALICE
AliEn Top Level Services	ALICE
UI(s) for submission to LCG/SC3	ALICE+LCG
File Catalog	ALICE
WMS (n RBs) + CE/SE Services on SC3	LCG
LFC instances on all SC3 sites, seen from WNs/VO-boxes	LCG
xrootd	LCG
gliteFTS accessible from all WNs/VO-boxes	LCG
SRM accessible from all xrootd workers/VO-boxes	LCG
SC3 resources for ALICE (Computing/Storage)	LCG
Appropriate JDL files for the different tasks	ALICE

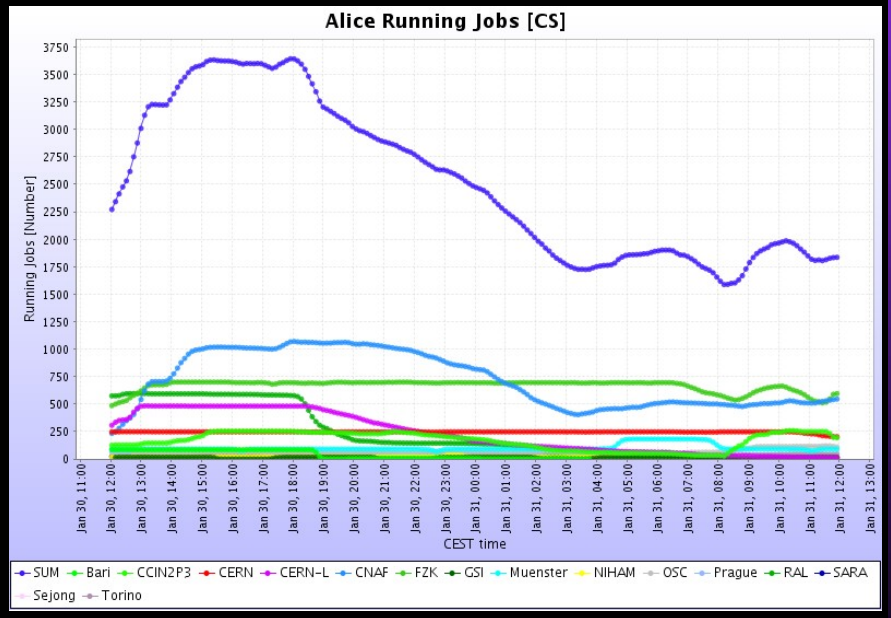
Schedule

- January 2006
  - Rerun of SC3 disk – disk transfers (max 150MB/s)
  - We should get ready to do this with the current data triggered via AliEn jobs or scheduled transfers
- March 2006
  - T0-T1 “loop-back” tests at 2 x nominal rate (CERN)
  - We prepare to run our bulk production
  - (We get ready with proof@caf)
- April 2006
  - T0-T1 disk-disk (nominal rates) disk-tape (50-75MB/s)
  - We run our bulk production and send data back to CERN
  - First chance to push out data, reconstruction at CERN
  - (First tests with proof@caf)
- July 2006
  - T0-T1 disk-tape (nominal rates)
  - T1-T1, T1-T2, T2-T1 and other rates TBD according to CTDRs
  - Second chance to push out the data
  - Reconstruction at CERN and remote centres
- September 2006
  - Scheduled analysis challenge
  - Unscheduled challenge (target T2’s?)

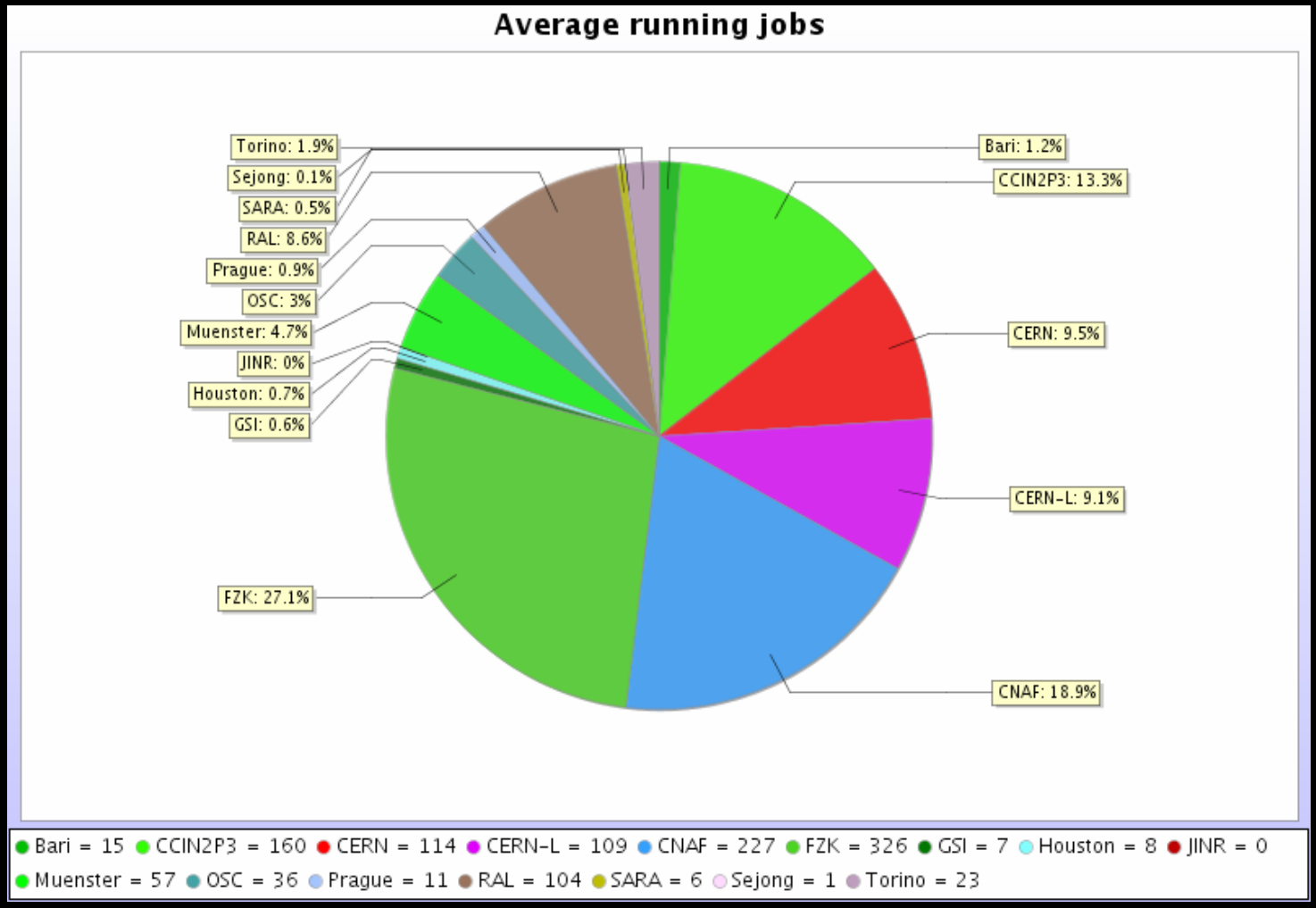
Data Challenge

- Test of the system started with simulation
- Up to 3600 jobs running in parallel
- Next will be reconstruction and analysis





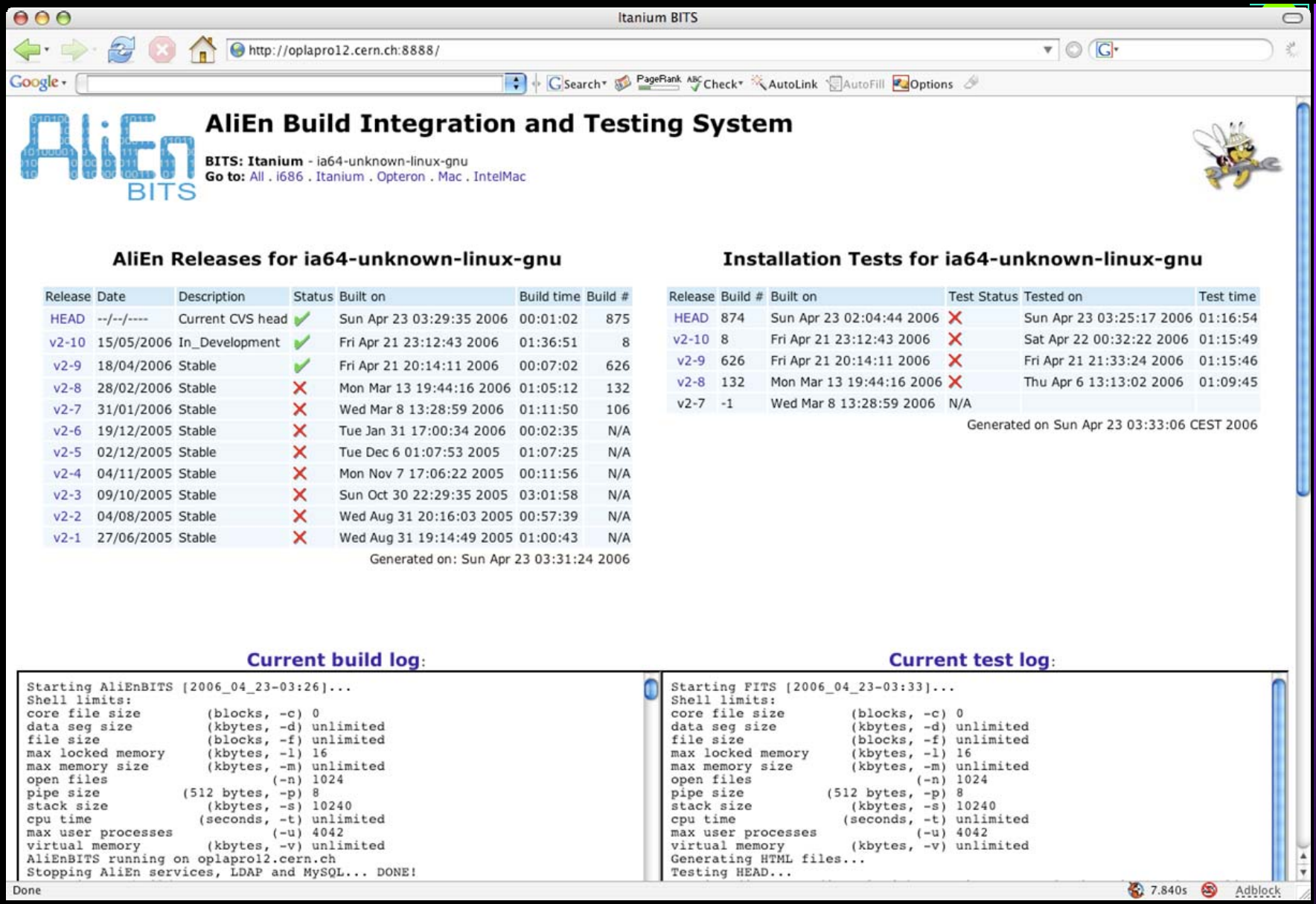
Statistics so far...



Finally,
What about the Itaniums performance?

(This is the second element of the Tragedy)

- Caveats:
  - Compilers are generally those available under GNU Linux.
  - Software design has not been tailored for individual platforms.
  - Software ported to: IA32, IA64, Opteron, Mac (PPC and mactel), SUN & even DEC.


Itanium BITS
http://oplapro12.cern.ch:8888/
AliEn Build Integration and Testing System
BITS: Itanium - ia64-unknown-linux-gnu
Go to: All . i686 . Itanium . Opteron . Mac . IntelMac
AliEn Releases for ia64-unknown-linux-gnu
Release Date Description Status Built on Build time Build #
HEAD --/--/---- Current CVS head Sun Apr 23 03:29:35 2006 00:01:02 875
v2-10 15/05/2006 In_Development Fri Apr 21 23:12:43 2006 01:36:51 8
v2-9 18/04/2006 Stable Fri Apr 21 20:14:11 2006 00:07:02 626
v2-8 28/02/2006 Stable Mon Mar 13 19:44:16 2006 01:05:12 132
v2-7 31/01/2006 Stable Wed Mar 8 13:28:59 2006 01:11:50 106
v2-6 19/12/2005 Stable Tue Jan 31 17:00:34 2006 00:02:35 N/A
v2-5 02/12/2005 Stable Tue Dec 6 01:07:53 2005 01:07:25 N/A
v2-4 04/11/2005 Stable Mon Nov 7 17:06:22 2005 00:11:56 N/A
v2-3 09/10/2005 Stable Sun Oct 30 22:29:35 2005 03:01:58 N/A
v2-2 04/08/2005 Stable Wed Aug 31 20:16:03 2005 00:57:39 N/A
v2-1 27/06/2005 Stable Wed Aug 31 19:14:49 2005 01:00:43 N/A
Generated on: Sun Apr 23 03:31:24 2006
Installation Tests for ia64-unknown-linux-gnu
Release Build # Built on Test Status Tested on Test time
HEAD 874 Sun Apr 23 02:04:44 2006 Sun Apr 23 03:25:17 2006 01:16:54
v2-10 8 Fri Apr 21 23:12:43 2006 Sat Apr 22 00:32:22 2006 01:15:49
v2-9 626 Fri Apr 21 20:14:11 2006 Fri Apr 21 21:33:24 2006 01:15:46
v2-8 132 Mon Mar 13 19:44:16 2006 Thu Apr 6 13:13:02 2006 01:09:45
v2-7 -1 Wed Mar 8 13:28:59 2006 N/A
Generated on Sun Apr 23 03:33:06 CEST 2006
Current build log:
Starting AliEnBITS [2006_04_23-03:26]...
Shell limits:
core file size (blocks, -c) 0
data seg size (kbytes, -d) unlimited
file size (blocks, -f) unlimited
max locked memory (kbytes, -l) 16
max memory size (kbytes, -m) unlimited
open files (-n) 1024
pipe size (512 bytes, -p) 8
stack size (kbytes, -s) 10240
cpu time (seconds, -t) unlimited
max user processes (-u) 4042
virtual memory (kbytes, -v) unlimited
AliEnBITS running on oplapro12.cern.ch
Stopping AliEn services, LDAP and MySQL... DONE!
Current test log:
Starting FITS [2006_04_23-03:33]...
Shell limits:
core file size (blocks, -c) 0
data seg size (kbytes, -d) unlimited
file size (blocks, -f) unlimited
max locked memory (kbytes, -l) 16
max memory size (kbytes, -m) unlimited
open files (-n) 1024
pipe size (512 bytes, -p) 8
stack size (kbytes, -s) 10240
cpu time (seconds, -t) unlimited
max user processes (-u) 4042
virtual memory (kbytes, -v) unlimited
Generating HTML files...
Testing HEAD...
Done
7.840s
Adblock

Current CPU Usage

- 90% IA32
- 2% IA64 (diminishing, was 5% in 2005, predominantly US -> Houston & OSC)
- 7% AMD (expected to grow rapidly, replaces IA32)
- 1% MAC PPC (g5's)

General Comments

- The software is not specifically optimized to run on a given platform.
- Rule of thumb is that all other parameters being equal (RAM especially), the execution time is proportional to the CPU frequency, therefore IA64 is the slowest.
- The central database services (backbone of the AliEn Grid) use Itaniums as DB machines (MySQL) HERE IS THE GOOD NEWS:
  - Stability is very good…
  - The DB is optimized for a 64 bit platform, therefore IA64 is very competitive…

Future Plans (The Final Piece of the Tragedy)

- Software: Continue to build and test on every platform available…
- Running: We do not (because we cannot) dictate the site fabric, however will use whatever is available. This year we aim at reaching up to 5000 jobs running in parallel on ~30 computing centers worldwide.
- HERE IS THE BAD NEWS:
  - Our experience shows that sites favor expanding with AMD-based (Opteron) worker-nodes (very favorable price/performance ratio).

The Tragedy Revisited...

- If we could tailor the software to favor the Itaniums, their performance might be competitive or better...
- Because we generally do not own our hardware, (The Genuine Grid Paradigm), we have less say in what it is...
- And, so far we have not had the manpower to worry about tailoring the software for different platforms, when we have to be able to run on everything we are offered...


"That's
all
folks!"

AliEn

- Coherent set of modular services
  - Used in production 2001-2004
- LCG elements have progressively replaced AliEn ones
  - Consistent with the plan announced since 2001 by ALICE
  - This will continue as suitable components become available
- Whenever possible, we use “common” services
- AliEn offers a single interface for ALICE users into the complex, heterogeneous (multiple grids and platforms) and fast-evolving Grid reality

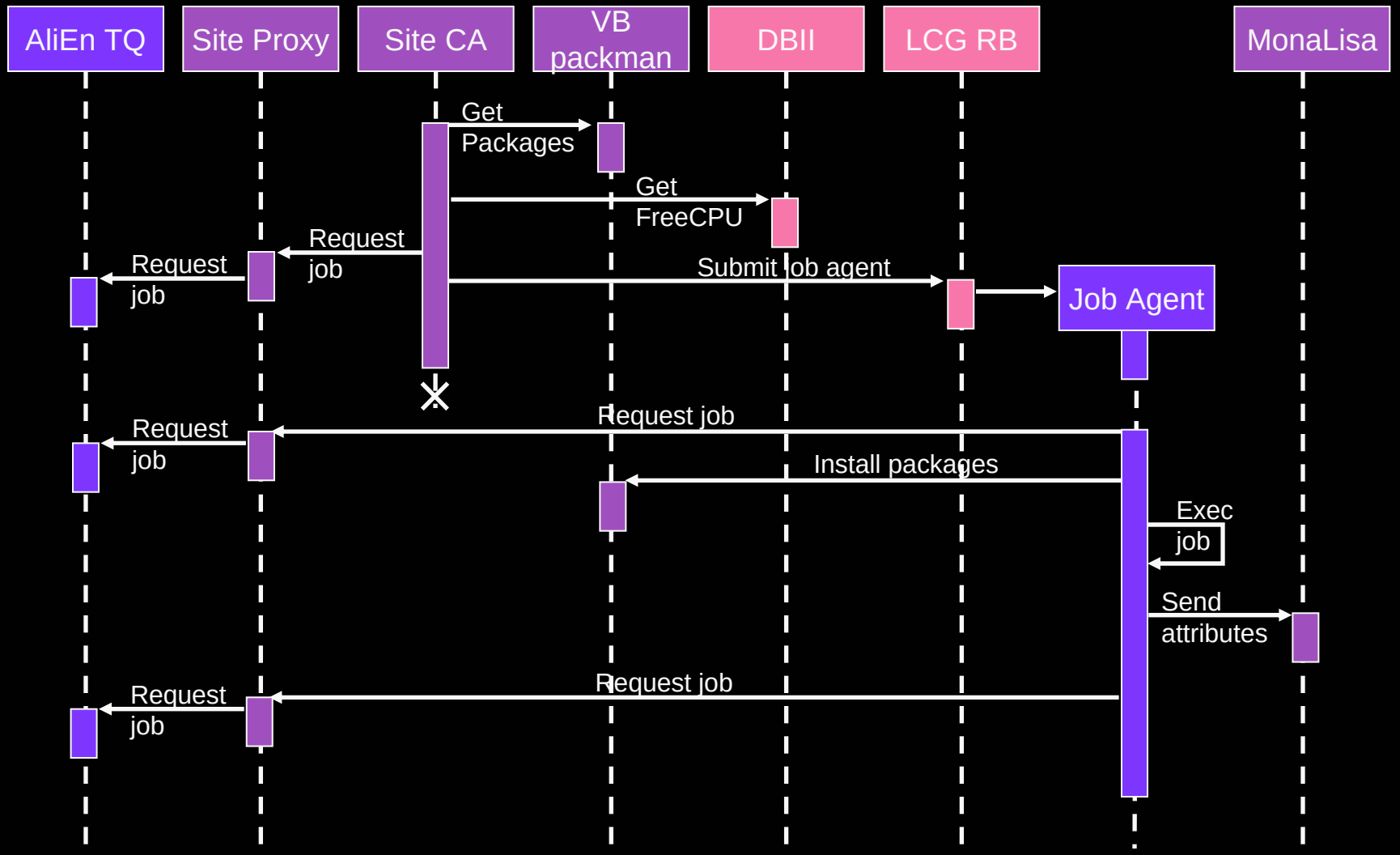
Services

- LCG services
  - LCG data management components (LFC, SRM)
  - Workload Management System (Resource Broker)
  - gLite Data Management components (FTS)
  - Virtual Organisation Management System (VOMS)
  - Common authentication model (GLOBUS)
  - Discovery service (planned)
- Discussed by the BS WG, coordinated by the ALICE-LCG-TF, tested in the DC
- AliEn services
  - ALICE job database and related distributed tools and services
  - ALICE file and dataset catalogue and related distributed tools and services
  - ALICE job reporting services
- Essential components for distributed data processing
  - Their functionality is ALICE-specific and not found elsewhere
  - They are integral part of the ALICE Computing Environment
- Interface with ARC is in progress, we are discussing with OSG

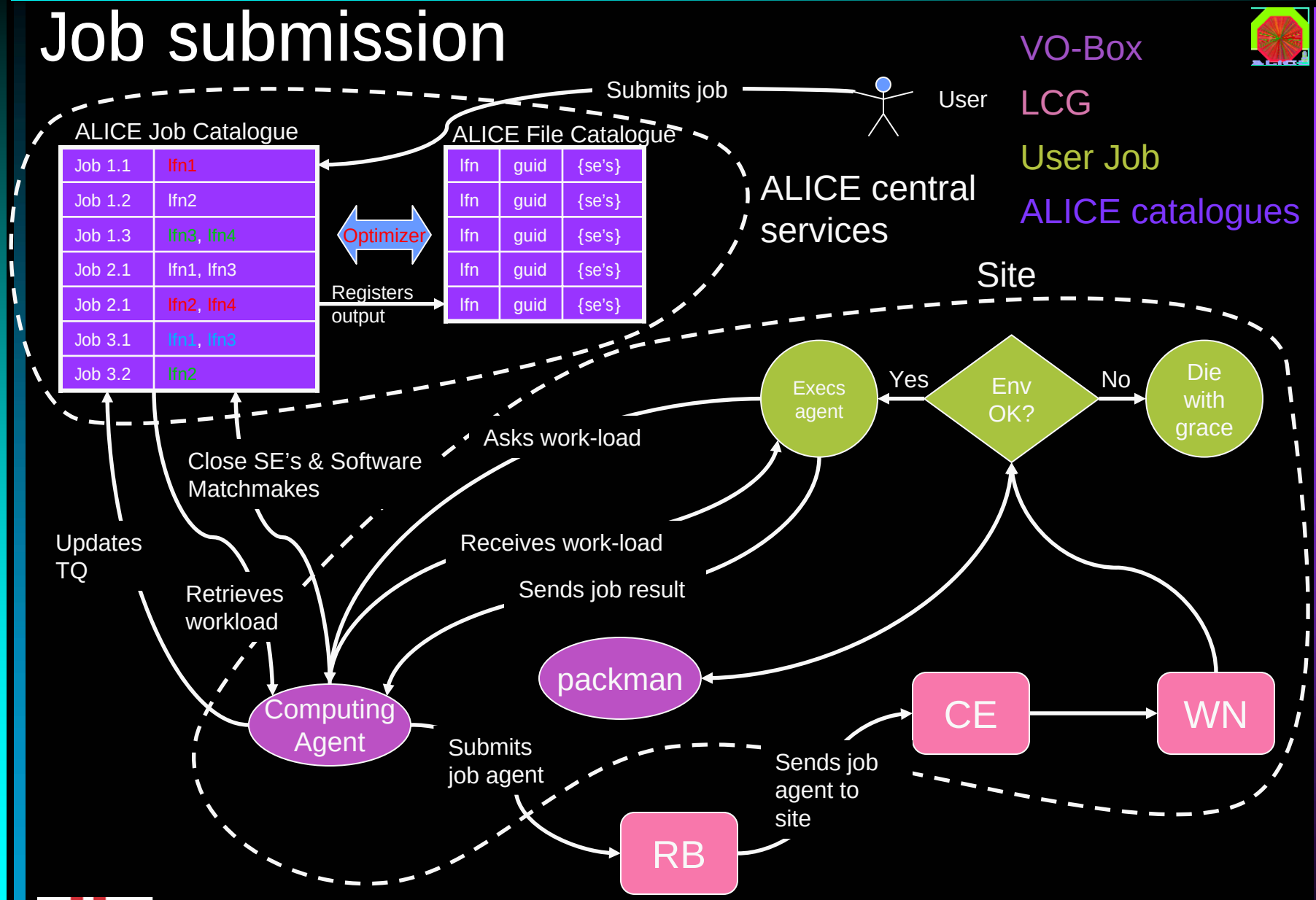
WMS interaction diagram

VO-Box User Job

ALICE catalogues LCG



Job submission



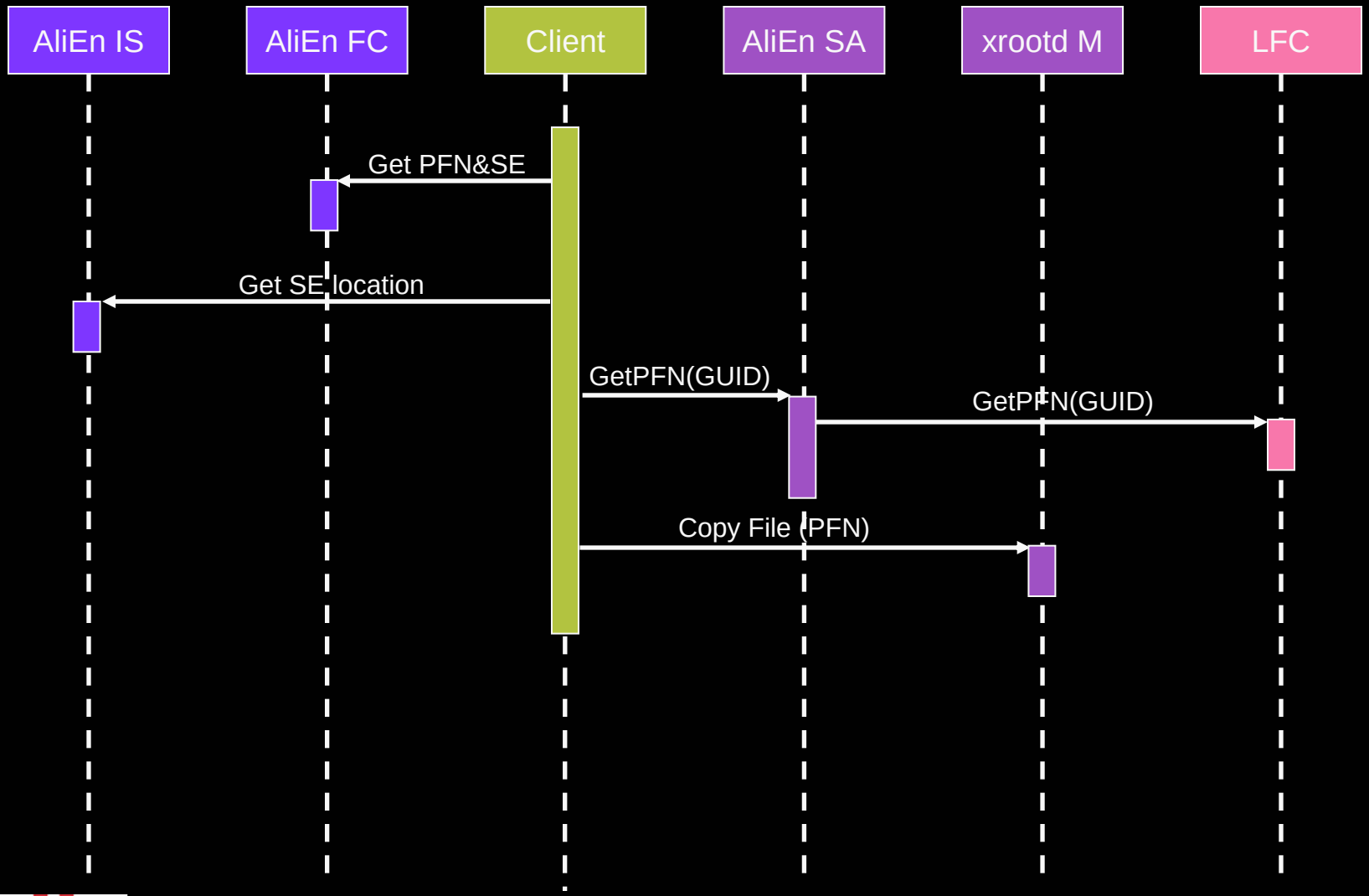
Job submission

- Job agents
  - Sent only when needed
  - Avoid waste of resources and “useless” updates of the ALICE Job Catalogue
  - Eliminate “black hole” effect
- Job location determined by the data location
- WN outbound connectivity required
  - We are working on removing this constraint
- System used for large production
  - 22,500 jobs, 540 KSi2K hours, 20TB
  - 2.5% inefficiency thanks to job agents

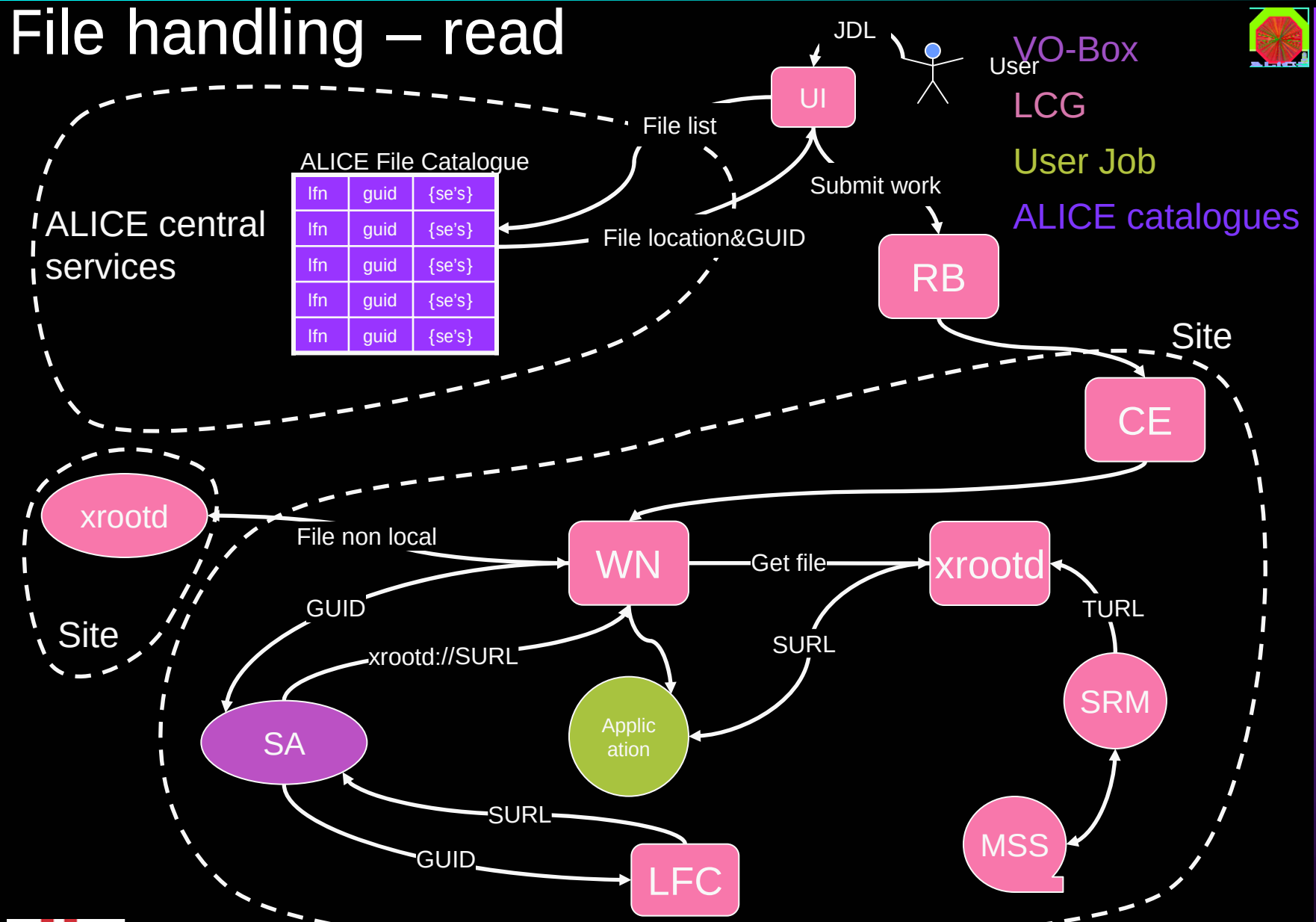
File retrieval

VO-Box User Job

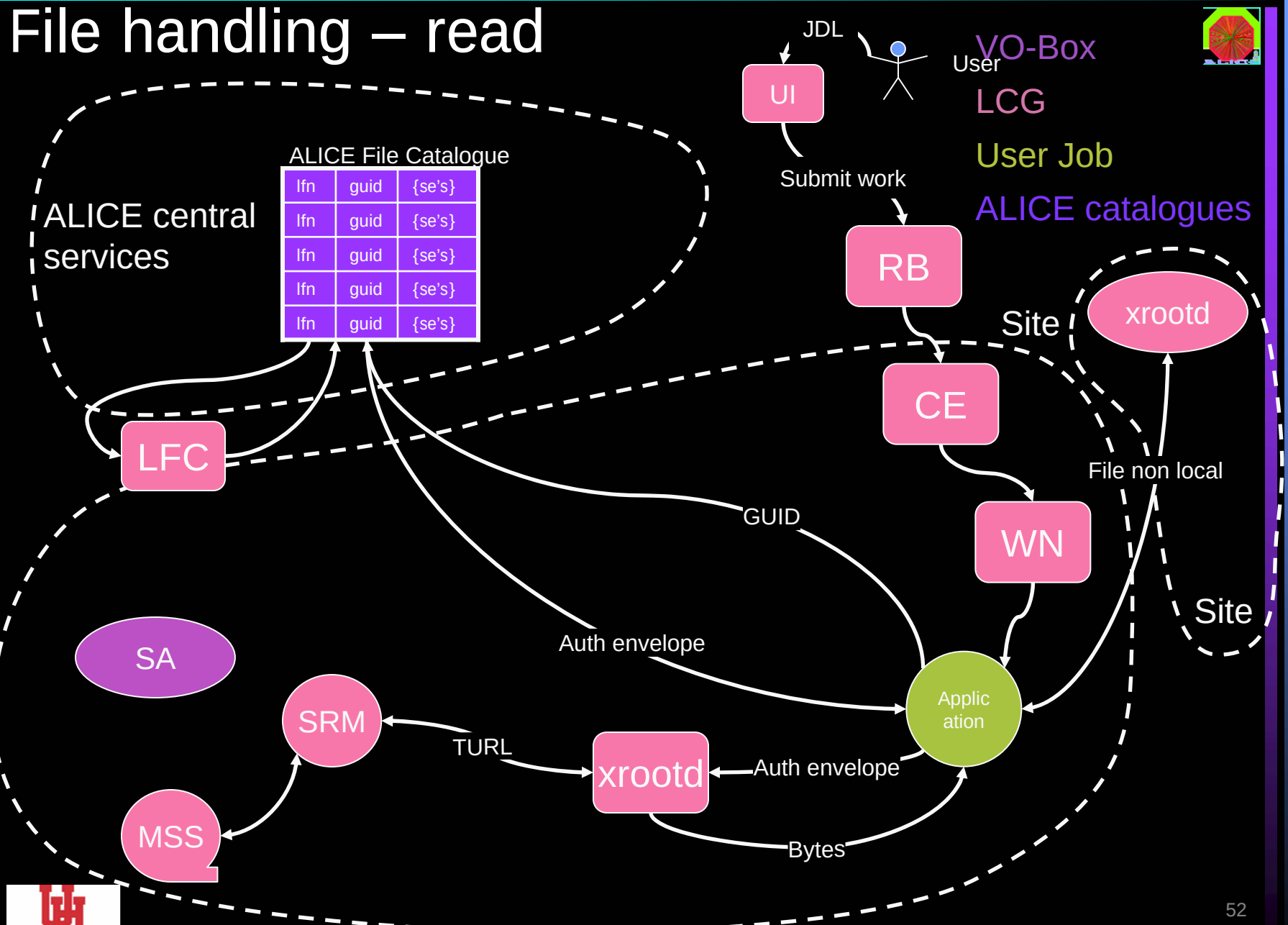
ALICE catalogues LCG



File handling – read



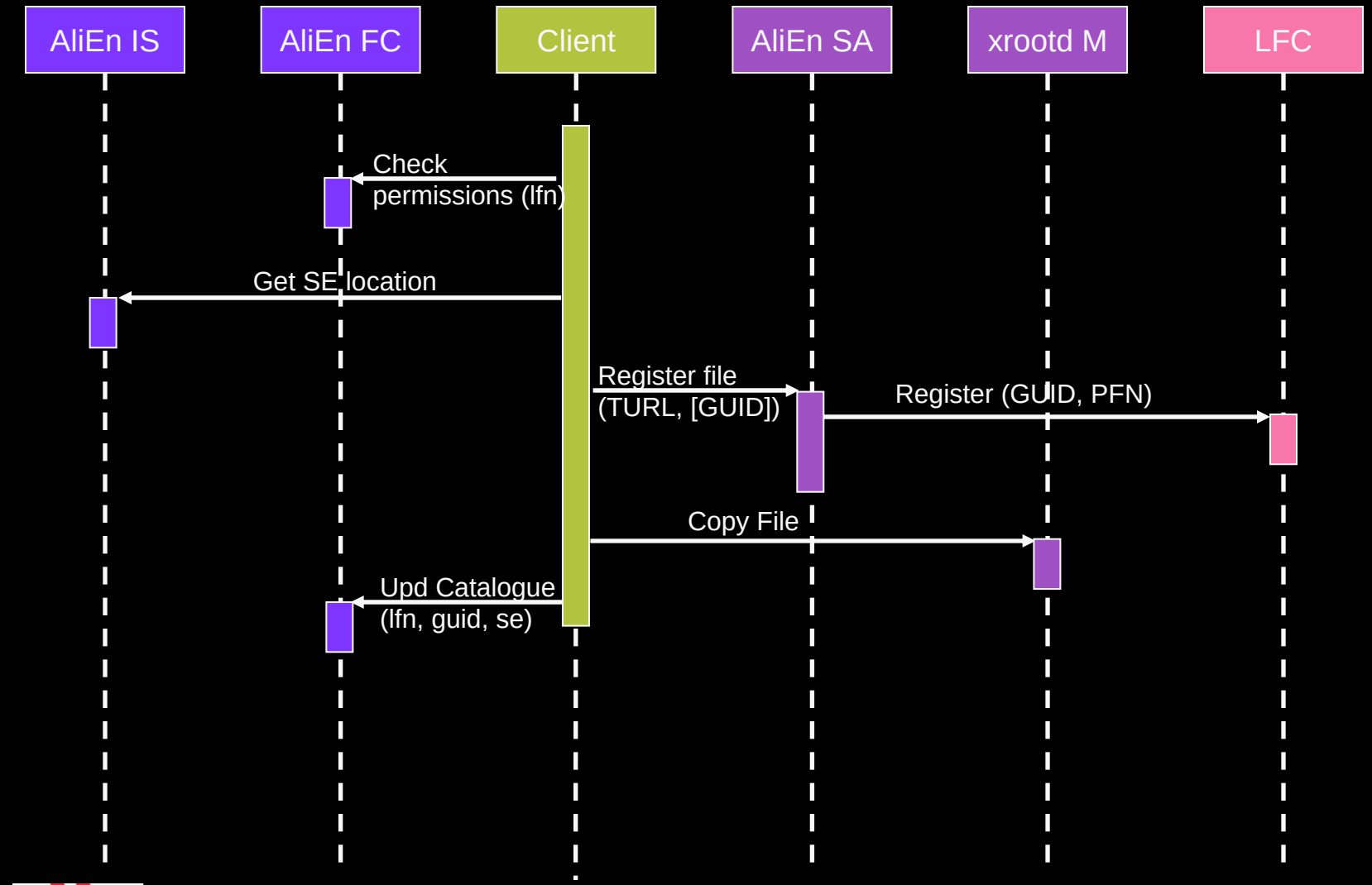
File handling – read



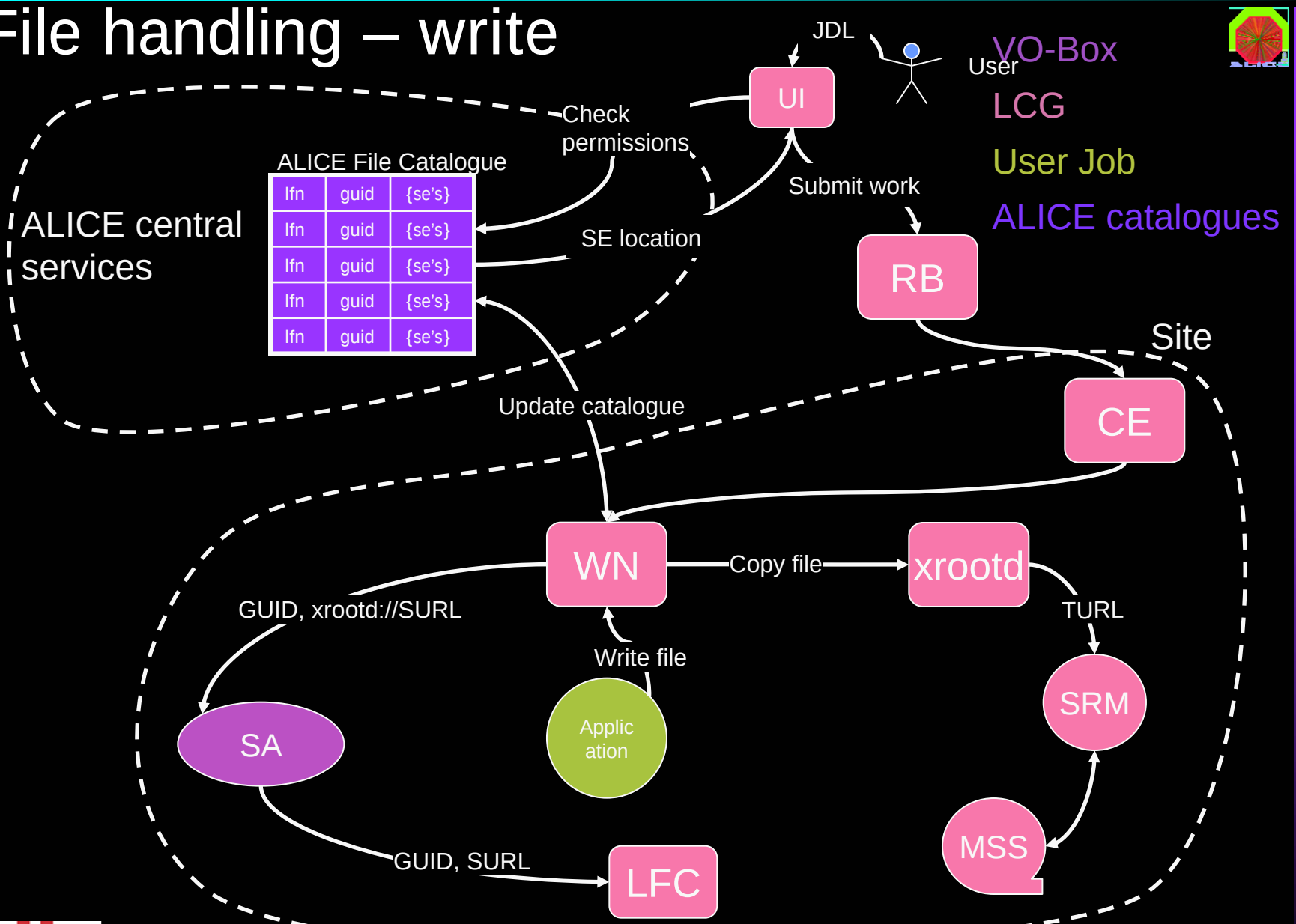
File registration

VO-Box User Job

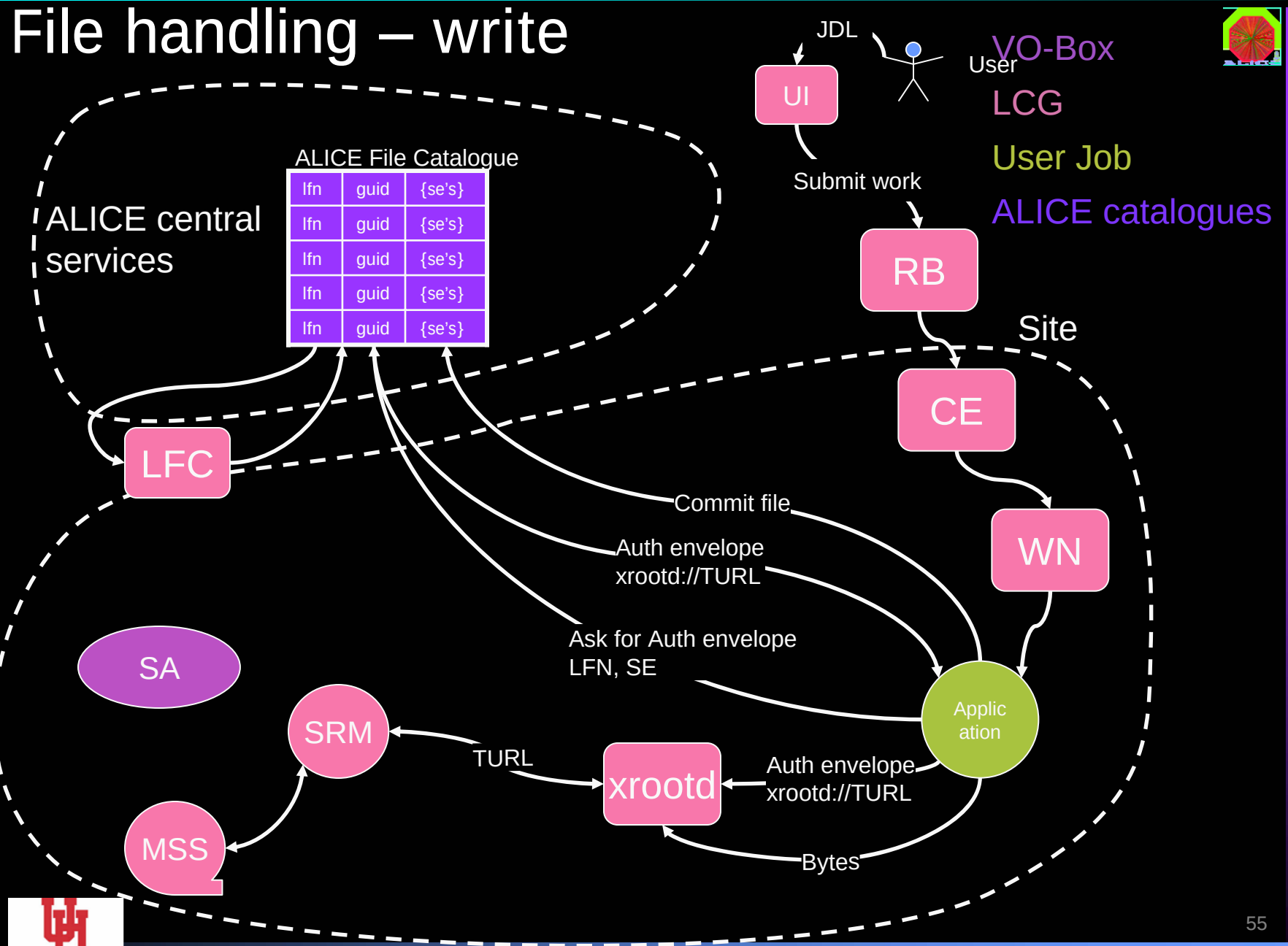
ALICE catalogues LCG



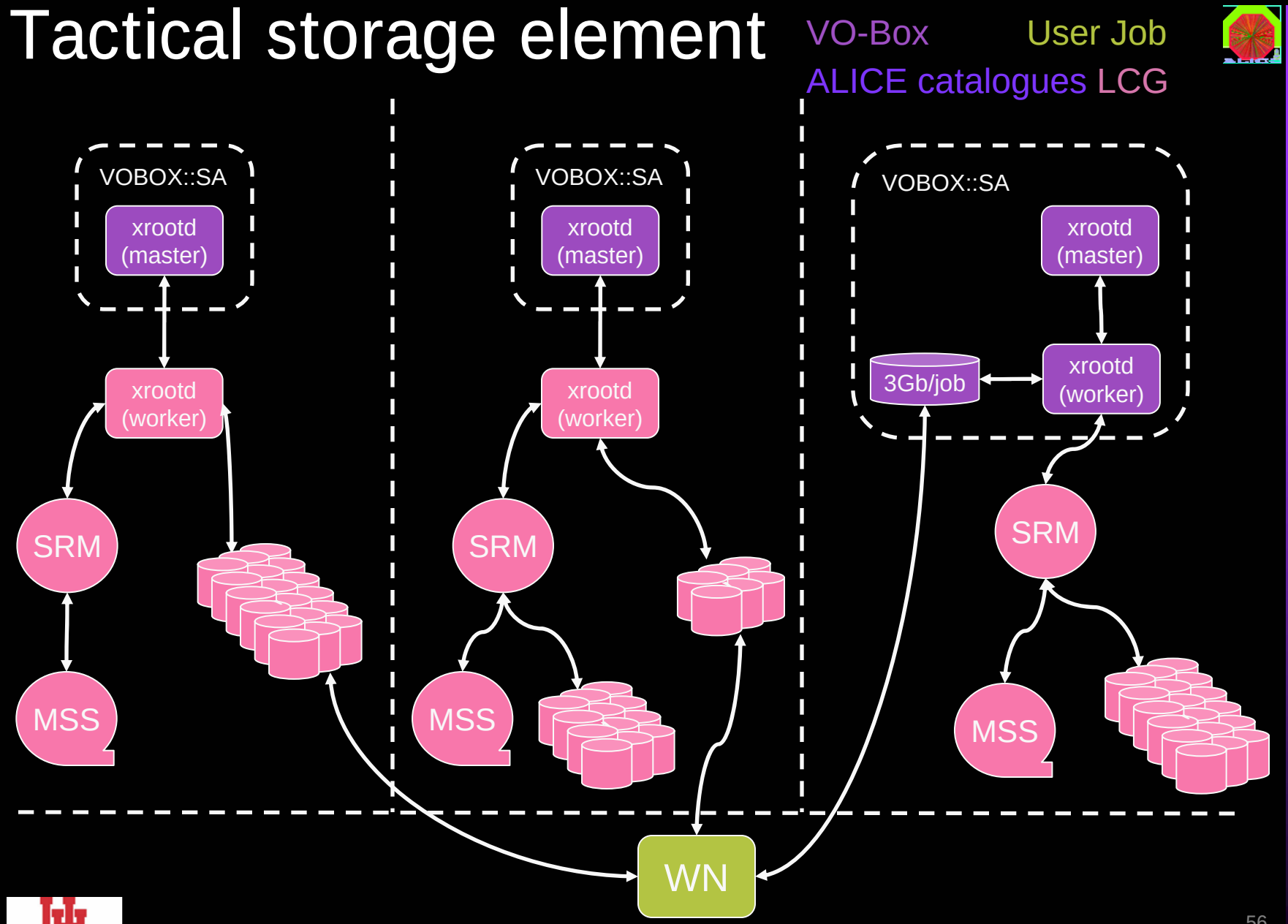
File handling – write



File handling – write



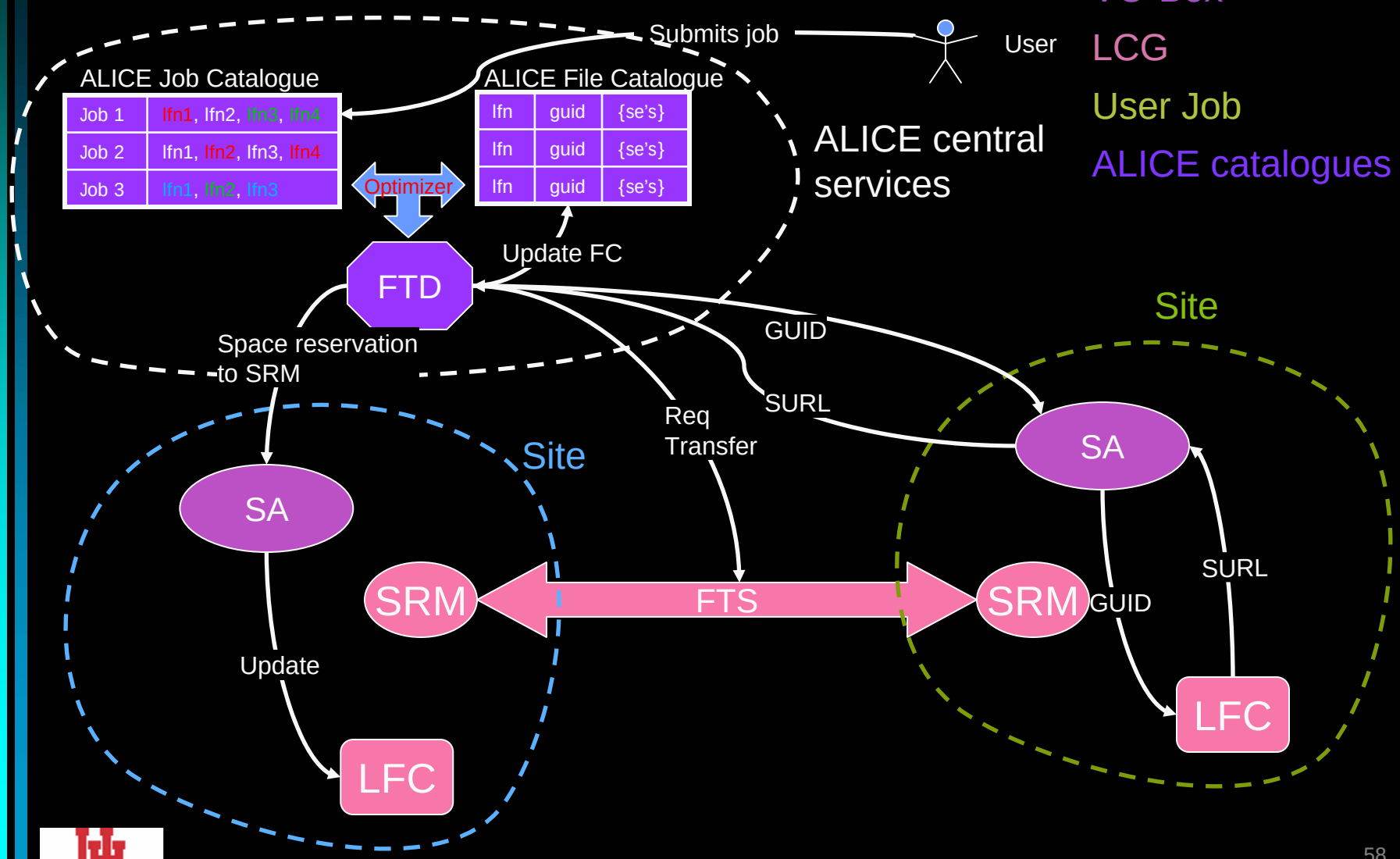
Tactical storage element



File handling

- ALICE File Catalogue contains ALICE MetaData
- Storage Adapter is our lightweight service which uses functionality of LFC to provide GUID⇒SURL translation
- User-level authorisation in the File Catalogue
- xrootd as a single protocol
- Simple extension to Grid Collector and proof
  - Already tested
- xrootd plugin will be part of the next xrootd release
- xrootd interface to SRM under discussion

File replication



LCG services we are using

- RB / WMS
- VOMS
- FTS / SRM
- LFC
- CASTOR2 (not LCG strictly speaking)
- Which is more or less all is there

LCG services we not are using

- In first approximation none
- However we have little / no experience with
  - dCACHE
  - DPM
- Which are not (again strictly speaking) LCG services

ALICE specific services

- Central services
  - Central TQ
  - File catalogue
  - User authentication using VOMS
  - FTD
- Site VO-Box services
  - PackMan
  - MonALISA
  - Site Computing Agent
  - Storage Adaptor
  - xrootd
  - Site proxy
  - proofd
  - Agent monitoring service

ALICE specific central services

- No real need to replace them, they come as a single instance
- Using standard components could reduce
  - The maintenance load
  - The hardware investment
- The following is largely an academic exercise

Central TQ

- Why needed
  - Bulk job submission
  - Job splitting based on data locality
  - Job data locality optimisation (with our FC!)
  - Job prioritisation within VO
  - Job bookkeeping with status and error codes specific for ALICE / AliRoot
- LCG equivalent
  - None
- How to replace
  - A generic central TQ as a BS service with the above functionality
  - This means “entre autre”
    - A plugin for our FC
    - Custom error and status codes
    - A plugin for our splitter / optimiser
    - ...

Central FC

- Why needed
  - ALICE specific MetaData (two years of development!)
  - Integrated with our TQ and optimiser
  - Integrated with our Grid Collector TAG MD
  - Highly optimised for GUID searches
  - User level file permission
- LCG equivalent
  - LFC
- How to replace
  - Major development to integrate it with all the above elements

User authentication using VOMS

- Why needed
  - To implement VOMS authentication in our central services
- LCG equivalent
  - N/A
- How to replace
  - Needed as long as we have our common services

FTD

- Why needed
  - To schedule and optimise transfers with FTS
  - Integrated with TQ and optimiser
  - Update our FC upon successful file transfer using FTS
- LCG equivalent
  - A module called FPS was foreseen
- How to replace
  - With a high-level service to schedule and optimise transfers with FTS
  - Integration with our optimiser and TQ
  - This service should have mechanism to update our file catalogue

ALICE specific site services

- Replacing all of them would mean to get rid of ALICE-maintained VO-Boxes
- The advantage is obvious in terms of maintenance load for ALICE

PackMan

- Why needed
  - Distribution, installation and configuration (env. variables etc.) of application software
  - Includes versioning and test tools
  - Integrated with TQ and matchmaking
  - A package is automatically installed if a job needs it before being “pulled” from the TQ
- LCG equivalent
  - Installation via normal jobs
- How to replace
  - Need development of a new tool
- Why on VO-Box
  - Need local access to shared software area

MonALISA

- Why needed
  - Monitoring of jobs, storage and traffic
  - ALICE specific monitoring
- LCG equivalent
  - RGMA, Grid-ICE
- How to replace
  - We would like to have MonALISA as a BS service (see Dashboard project)
- Why on VO-Box
  - Local aggregation minimising monitoring traffic
  - Monitor of the VO-Box itself

Site Computing Agent

- Why needed
  - Interfaces to WMS (LCG, ARC, OSG…)
  - *Matchmakes* within ALICE TQ
- LCG equivalent
  - N/A
- How to replace
  - N/A
- Why on VO-Box
  - Can be done centrally
  - Problems of scaling and size of the CERN ALICE installation

Storage adaptor

- Why needed
  - Handles communication with LFC to translate GUID to TURL/SURL
  - Builds the TURL in case of write
  - Can act as a volume manager
  - Starts up xrootd services
  - Handles communication with FTD
  - Monitor site storage configuration
- LCG equivalent
  - N/A
- How to replace
  - N/A
- Why on VO-Box
  - Avoid communication with a central service (see LHCb experience)
  - Need to communicate with local LFC
  - Need to be local to start xrootd and to monitor the storage

xrootd

- Why needed
  - Allows posix I/O
  - Insulates application from local storage systems (different I/O libraries)
  - Efficient handling of storage
  - Handles user-level file authorisation
- LCG equivalent
  - Nothing providing all of the above
- How to replace
  - Should be introduced as BS service
- Why on VO-Box
  - Need to communicate with local storage system

Site proxy

- Why needed
  - Communication between job agent and central services in case WN do not have incoming / outgoing connectivity
- LCG equivalent
  - N/A
- How to replace
  - Needed as long as we have ALICE central services
- Why on VO-Box
  - Need to communicate with local services
  - Essential to handle local VO-Specific services

proofd

- Why needed
  - To handle communication between local master and workers when using proof on wide area
- LCG equivalent
  - N/A
- How to replace
  - proof should be a BS service
- Why on VO-Box
  - Need to communicate with local WN's
  - This service is now part of xrootd

Agent monitoring service

- Why needed
  - Heart-beat for the VO-Box
- LCG equivalent
  - N/A
- How to replace
  - Needed as long as we have ALICE VO-Boxes
- Why on VO-Box
  - Obvious...

Possible developments

- xrootd can also open “remote” files
  - Application can open files not local and not declared in the JDL
  - Application can read bits of files instead than transferring them completely
  - Can have an “extended” list of close SE’s
  - Can be an excellent complement to job splitting
- Try to reduce the need for WN outbound connectivity
  - More acceptable to sysadmin
  - More proxying implied

ALICE VO-Box requirements

- General requirements:
  - PIII 2GHz, 1024 MB RAM. Any Linux flavour, kernel 2.4+.
  - User accounts for SGM's, via gsissh
  - UI fully implemented, lcg-infosites and FTS
  - Access to the experiment software installation area
- Agents and services
  - Site service interfaces and monitoring agents:
    - Storage Adapter (SA), File Transfer Daemon (Interface to FTS)
    - Site Proxy (SP), MonALISA, Agents Monitoring
    - Site Computing Agent (Interface to LCG RB)
    - PackMan (PM), xrootd
    - Agent Monitoring (AmOn)
    - proofd daemon for interactive distributed analysis with PROOF
  - Connectivity
    - Outbound connectivity + Access to local storage (direct or SRM)
    - Inbound connectivity on some fixed network ports
      - From CERN, for SP and PM (e.g.: 8084 and 9991)
      - From World, for SA and xrootd (e.g.: 8082 and 1094)
      - From CERN for proofd (e.g.: 1093)

Operational conditions

- We have no problem with the operational holy services
  - Workload will be accepted ONLY through the standard gatekeeper
    - Yes!
  - Data for storage will be accepted ONLY through the standard SRM interface
    - Yes!
  - If the UI is required on the VOBox, it must be the LCG UI
    - Yes!
- Can provide at any moment the list of all programs running
- Do not require root account
- A fixed number of ports with well defined connectivity
- Any addition / change / removal of code can be discussed with the appropriate bodies


"That's
all
folks!"